CHAPTER 11

Stage 1
Problem definition

Stage 2
Research approach developed

Stage 3
Research design developed

Stage 4
Fieldwork or data collection

Stage 5
Data preparation and analysis

Stage 6
Report preparation and presentation

Causal research design: experimentation

Objectives

After reading this chapter, you should be able to:

1. explain the concept of causality as defined in marketing research and distinguish between the ordinary meaning and the scientific meaning of causality;
2. define and differentiate two types of validity: internal validity and external validity;
3. discuss the various extraneous variables that can affect the validity of results obtained through experimentation and explain how the researcher can control extraneous variables;
4. describe and evaluate experimental designs and the differences among pre-experimental, true experimental, quasi-experimental and statistical designs;
5. compare and contrast the use of laboratory versus field experimentation and experimental versus non-experimental designs in marketing research;
6. describe test marketing and its various forms: standard test market, controlled test market and simulated test market;
7. understand the problems of internal and external validity of field experiments when conducted in international markets;
8. describe the ethical issues involved in conducting causal research and the role of debriefing in addressing some of these issues.

Causality can never be proved; in other words, it can never be demonstrated decisively. Inferences of cause-and-effect relationships are the best that can be achieved.

Overview

We introduced causal designs in Chapter 3, where we discussed their relationship to exploratory and descriptive designs and defined experimentation as the primary method employed in causal designs. This chapter explores the concept of causality further. We identify the necessary conditions for causality, examine the role of validity in experimentation, and consider the extraneous variables and procedures for controlling them. We present a classification of experimental designs and consider specific designs, along with the relative merits of laboratory and field experiments. An application in the area of test marketing is discussed in detail. The considerations involved in conducting experimental research when researching international markets are discussed. Several ethical issues, which arise in experimentation, are identified. We begin with an example that encapsulates the application and process of experimentation.

example

POP buys[1]

The Eckerd Drug Company conducted an experiment to examine the effectiveness of in-store radio advertisements to induce point-of-purchase (POP) buys. Twenty statistically compatible stores were selected based on store size, geographical location, traffic flow count and age. Half of these were randomly selected as test stores, and the other half served as control stores. The test stores aired the radio advertisements, whereas the control stores' POP radio systems were removed. Tracking data in the form of unit sales and turnover were obtained for the following three periods: seven days before the experiment, during the course of the four-week experiment, and seven days after the experiment. The products monitored varied from inexpensive items to small kitchen appliances. Results indicated that sales of the advertised products in the test stores at least doubled. Based on this evidence, Eckerd concluded that in-store radio advertising was highly effective in inducing POP buys, and they decided to continue it. ■

Concept of causality

Causality
Causality applies when the occurrence of X increases the probability of the occurrence of Y.

Experimentation is commonly used to infer causal relationships. The concept of **causality** requires some explanation. The scientific concept of causality is complex. 'Causality' means something very different to the average person on the street than to a scientist.[2] A statement such as 'X causes Y' will have the following meaning to an ordinary person and to a scientist.

Ordinary meaning	*Scientific meaning*
X is the only cause of Y.	X is only one of a number of possible causes of Y.
X must always lead to Y.	The occurrence of X makes the occurrence of Y more probable (X is a probabilistic cause of Y).
It is possible to prove that X is a cause of Y.	We can never prove that X is a cause of Y. At best, we can infer that X is a cause of Y.

The scientific meaning of causality is more appropriate to marketing research than is the everyday meaning.[3] Marketing effects are caused by multiple variables and the relationship between cause and effect tends to be probabilistic. Moreover, we can never prove causality (i.e. demonstrate it conclusively); we can only infer a cause-and-effect relationship. In other words, it is possible that the true causal relation, if one exists, will not have been identified. We further clarify the concept of causality by discussing the conditions for causality.

Conditions for causality

Before making causal inferences, or assuming causality, three conditions must be satisfied: (1) concomitant variation, (2) time order of occurrence of variables, and (3) elimination of other possible causal factors. These conditions are necessary but not sufficient to demonstrate causality. No one of these three conditions, nor all three conditions combined, can demonstrate decisively that a causal relationship exists.[4] These conditions are explained in more detail in the following sections.

Concomitant variation

Concomitant variation
A condition for inferring causality that requires that the extent to which a cause, X, and an effect, Y, occur together or vary together is predicted by the hypothesis under consideration.

Concomitant variation is the extent to which a cause, X, and an effect, Y, occur together or vary together in the way predicted by the hypothesis under consideration. Evidence pertaining to concomitant variation can be obtained in a qualitative or quantitative manner.

For example, in the qualitative case, the management of a bank may believe that the retention of customers is highly dependent on the quality of service in bank branches. This hypothesis could be examined by assessing concomitant variation. Here, the causal factor X is branch service and the effect factor Y is retention level. A concomitant variation supporting the hypothesis would imply that banks with satisfactory levels of service would also have a satisfactory retention of customers. Likewise, banks with unsatisfactory service would exhibit unsatisfactory retention of customers. If, on the other hand, the opposite pattern was found, we would conclude that the hypothesis was untenable.

For a quantitative example, consider a random survey of 1,000 respondents questioned on the purchase of shares from a bank branch. This survey yields the data in Table 11.1. The respondents have been classified into high- and low-education groups based on a median or even split. This table suggests that the purchase of shares is influenced by education level. Respondents with high education are likely to purchase more shares. Seventy-three per cent of the respondents with high education have a high purchase level, whereas only 64% of those with low education have a high purchase level. Furthermore, this is based on a relatively large sample of 1,000 people.

Table 11.1 **Evidence of concomitant variation between purchase of shares and education**

		Purchase of shares from a bank, Y		
		High	*Low*	*Total*
Education, X	*High*	363 (73%)	137 (27%)	500 (100%)
	Low	322 (64%)	178 (36%)	500 (100%)

Based on this evidence, can we conclude that high education causes high purchase of shares? Certainly not! All that can be said is that association makes the hypothesis more tenable; it does not prove it. What about the effect of other possible causal factors such as income? Shares can be expensive, so people with higher incomes may buy more of them. Table 11.2 shows the relationship between purchase of shares and education for different income segments. This is equivalent to holding the effect of income constant. Here again, the sample has been split at the median to produce high- and low-income groups of equal size. Table 11.2 shows that the difference in purchase of shares between high- and low-education respondents has been reduced considerably. This suggests that the association indicated by Table 11.1 may be spurious.

Table 11.2 Purchase of shares by income and education

		Low-income purchase		
		High	*Low*	*Total*
Education	*High*	122 (61%)	78 (39%)	200 (100%)
	Low	171 (57%)	129 (43%)	300 (100%)

		High-income purchase		
		High	*Low*	*Total*
Education	*High*	241 (80%)	59 (20%)	300 (100%)
	Low	151 (76%)	49 (24%)	200 (100%)

We could give similar examples to show why the absence of initial evidence of concomitant variation does not imply that there is no causation. It is possible that considering a third variable will crystallise an association that was originally obscure. The time order of the occurrence of variables provides additional insights into causality.

Time order of occurrence of variables

The time order of occurrence condition states that the causing event must occur either before or simultaneously with the effect; it cannot occur afterwards. By definition, an effect cannot be produced by an event that occurs after the effect has taken place. It is possible, however, for each event in a relationship to be both a cause and an effect of the other event. In other words, a variable can be both a cause and an effect in the same causal relationship. To illustrate, customers who shop frequently in a particular supermarket are more likely to have a loyalty card for that supermarket. In addition, customers who have a loyalty card for a supermarket are likely to shop there frequently.

Consider banks and the retention of customers. If in-bank service is the cause of retention, then improvements in service must be made before, or at least simultaneously with, an increase in retention. These improvements might consist of training or hiring more counter staff. Then, in subsequent months, the retention of bank customers should increase. Alternatively, retention may increase simultaneously with the training or hiring of additional counter staff. On the other hand, suppose that a bank experienced an appreciable increase in the level of retaining customers and then decided to use some of that money generated to retrain its counter staff, leading to an improvement in service. In this case, bank service cannot be a cause of increased retention; rather, just the opposite hypothesis might be plausible.

Elimination of other possible causal factors

The absence of other possible causal factors means that the factor or variable being investigated should be the only possible causal explanation. Bank service may be a cause of retention if we can be sure that changes in all other factors affecting retention: pricing, advertising, promotional offers, product characteristics, competition and so forth, were held constant or were otherwise controlled.

In an after-the-fact examination of a situation, we can never confidently rule out all other causal factors. In contrast, with experimental designs it is possible to control some of the other causal factors. It is also possible to balance the effects of some of the uncontrolled variables so that only random variations resulting from these uncontrolled variables will be measured. These aspects are discussed in more detail later in this chapter.

The difficulty of establishing a causal relationship is illustrated by the following example.

example

Which comes first?[5]

There are studies that contend that consumers increasingly make buying decisions in the store while they are shopping. Some studies indicate that as much as 80% of buying decisions are made at point-of-purchase (POP). POP buying decisions have increased concurrently with increased advertising efforts in the stores. These include radio advertisements, ads on shopping trolleys and grocery bags, ceiling signs and shelf displays. It is difficult to ascertain from these data whether the increased POP decision-making is the result of increased advertising efforts in the store or whether the increase in store advertising results from attempts to capture changing consumer attitudes towards purchasing and to capture sales from the increase in POP decision-making. It is also possible that both variables may be both causes and effects in this relationship. ■

If, as this example indicates, it is difficult to establish cause-and-effect relationships, what is the role of evidence obtained in experimentation? Evidence of concomitant variation, time order of occurrence of variables, and elimination of other possible causal factors, even if combined, still does not demonstrate conclusively that a causal relationship exists. If all the evidence is strong and consistent, however, it may be reasonable to conclude that there is a causal relationship. Accumulated evidence from several investigations increases our confidence that a causal relationship exists. Confidence is further enhanced if the evidence is interpreted in light of intimate conceptual knowledge of the problem situation. Controlled experiments can provide strong evidence on all three conditions.

Definitions and concepts

In this section, we define some basic concepts and illustrate them using examples.

Independent variables
Variables that are manipulated by the researcher and whose effects are measured and compared.

Test units
Individuals, organisations or other entities whose responses to independent variables or treatments are being studied.

Dependent variables
Variables that measure the effect of the independent variables on the test units.

Extraneous variables
Variables, other than dependent and independent variables, which may influence the results of the experiment.

Experiment
The process of manipulating one or more independent variables and measuring their effect on one or more dependent variables, while controlling for the extraneous variables.

- **Independent variables**. Independent variables are variables or alternatives that are manipulated (i.e. the levels of these variables are changed by the researcher) and whose effects are measured and compared. These variables, also known as treatments, may include price levels, package designs and advertising themes. In the Eckerd Drug Company example at the beginning of this chapter, the independent variable was 'in-store radio advertising' (present versus absent).
- **Test units**. Test units are individuals, organisations or other entities whose response to the independent variables or treatments is being examined. Test units may include consumers, stores or geographical areas. The test units were stores in the Eckerd example.
- **Dependent variables**. Dependent variables are the variables that measure the effect of the independent variables on the test units. These variables may include sales, profits and market shares. In the Eckerd example, the dependent variable was the sales level of advertised point-of-purchase products.
- **Extraneous variables**. Extraneous variables are all variables other than the independent variables that affect the response of the test units. These variables can confound the dependent variable measures in a way that weakens or invalidates the results of the experiment. In the Eckerd example, store size, geographical location, traffic flow count and age of the stores were extraneous variables that had to be controlled.
- **Experiment**. An experiment is formed when the researcher manipulates one or more independent variables and measures their effect on one or more dependent variables, while controlling for the effect of extraneous variables.[6] The Eckerd research project qualifies as an experiment based on this definition.

Experimental design
The set of experimental procedures specifying (1) the test units and sampling procedures, (2) the independent variables, (3) the dependent variables, and (4) how to control the extraneous variables.

- **Experimental design**. An experimental design is a set of procedures specifying: (1) the test units and how these units are to be divided into homogeneous sub-samples, (2) what independent variables or treatments are to be manipulated, (3) what dependent variables are to be measured, and (4) how the extraneous variables are to be controlled.[7]

As a further illustration of these definitions, consider the following example.

example

Taking coupons at face value[8]

An experiment was conducted to test the effects of the face value of sales promotion coupons (i.e. the amount saved when a consumer next buys the product) on the likelihood of consumers redeeming those coupons, controlling for the frequency of brand usage. Personal interviews were conducted with 280 shoppers who were entering or leaving a supermarket. Subjects were randomly assigned to two treatment groups. One offered low-value coupons and the other high-value coupons for four products: Tide detergent, Kellogg's Cornflakes, Aim toothpaste, and Joy liquid detergent. During the interviews, the respondents answered questions about which brands they used and how likely they were to cash the coupons of the given face value the next time they shopped. An interesting finding was that higher face-value coupons produced a greater likelihood of redemption among infrequent or non-buyers of the promoted brand but had little effect on regular buyers. ■

In the preceding experiment, the independent variable that was manipulated was the value of the coupon. The dependent variable was the likelihood of cashing the coupon. The extraneous variable that was controlled was brand usage. The test units were individual shoppers. The experimental design required the random assignment of test units (shoppers) to treatment groups (low or high value coupon).

Definition of symbols

To facilitate our discussion of extraneous variables and specific experimental designs, we define a set of symbols now commonly used in marketing research.[9]

X = the exposure of a group to an independent variable, treatment or event, the effects of which are to be determined
O = the process of observation or measurement of the dependent variable on the test units or group of units
R = the random assignment of test units or groups to separate treatments

In addition, the following conventions are adopted:

- Movement from left to right indicates movement through time.
- Horizontal alignment of symbols implies that all those symbols refer to a specific treatment group.
- Vertical alignment of symbols implies that those symbols refer to activities or events that occur simultaneously.

For example, the symbolic arrangement

$$\mathbf{X} \quad \mathbf{O}_1 \quad \mathbf{O}_2$$

means that a given group of test units was exposed to the treatment variable (**X**) and the response was measured at two different points in time $\mathbf{O}_1$ and $\mathbf{O}_2$.

Likewise, the symbolic arrangement

$$\begin{array}{lll} R & X_1 & O_1 \\ R & X_2 & O_2 \end{array}$$

means that two groups of test units were randomly assigned to two different treatment groups at the same time, and the dependent variable was measured in the two groups simultaneously.

Validity in experimentation

When conducting an experiment, a researcher has two goals: (1) to draw valid conclusions about the effects of independent variables on the study group, and (2) to make valid generalisations to a larger population of interest. The first goal concerns internal validity, the second external validity.[10]

Internal validity

Internal validity
A measure of accuracy of an experiment. It measures whether the manipulation of the independent variables, or treatments, actually caused the effects on the dependent variable(s).

Internal validity refers to whether the manipulation of the independent variables or treatments actually caused the observed effects on the dependent variables. Thus, internal validity refers to whether the observed effects on the test units could have been caused by variables other than the treatment. If the observed effects are influenced or confounded by extraneous variables, it is difficult to draw valid inferences about the causal relationship between the independent and dependent variables. Internal validity is the basic minimum that must be present in an experiment before any conclusion about treatment effects can be made. Without internal validity, the experimental results are confounded. Control of extraneous variables is a necessary condition for establishing internal validity.

External validity

External validity
A determination of whether the cause-and-effect relationships found in the experiment can be generalised.

External validity refers to whether the cause-and-effect relationships found in the experiment can be generalised. In other words, can the results be generalised beyond the experimental situation, and if so, to what populations, settings, times, independent variables and dependent variables can the results be projected?[11] Threats to external validity arise when the specific set of experimental conditions does not realistically take into account the interactions of other relevant variables in the real world.

It is desirable to have an experimental design that has both internal and external validity, but in applied marketing research we often have to trade one type of validity for another.[12] To control for extraneous variables, a researcher may conduct an experiment in an artificial environment. This enhances internal validity, but it may limit the generalisability of the results, thereby reducing external validity. For example, fast-food chains test customers' preferences for new formulations of menu items in test kitchens. Can the effects measured in this environment be generalised to fast-food outlets that may operate in multifarious environments? (Further discussion on the influence of artificiality on external validity may be found in the section of this chapter on laboratory versus field experimentation.) Regardless of the deterrents to external validity, if an experiment lacks internal validity, it may not be meaningful to generalise the results. Factors that threaten internal validity may also threaten external validity, the most serious of these being extraneous variables.

Extraneous variables

The need to control extraneous variables to establish internal and external validity has already been discussed. In this section, we classify extraneous variables in the following categories: history, maturation, testing effects, instrumentation, statistical regression, selection bias and mortality.

History

History
Specific events that are external to the experiment but that occur at the same time as the experiment.

Contrary to what the name implies, **history** (H) does not refer to the occurrence of events before the experiment. Rather, history refers to specific events that are external to the experiment but that occur at the same time as the experiment. These events may affect the dependent variable. Consider the following experiment:

$$\mathbf{O_1} \quad \mathbf{X_1} \quad \mathbf{O_2}$$

where $\mathbf{O_1}$ and $\mathbf{O_2}$ are measures of personal loan applications in a specific region and $\mathbf{X_1}$ represents a new promotional campaign. The difference $(\mathbf{O_2} - \mathbf{O_1})$ is the treatment effect. Suppose that the experiment revealed that there was no difference between $\mathbf{O_2}$ and $\mathbf{O_1}$. Can we then conclude that the promotional campaign was ineffective? Certainly not! The promotional campaign $\mathbf{X_1}$ is not the only possible explanation of the difference between $\mathbf{O_2}$ and $\mathbf{O_1}$. The campaign might well have been effective. What if general economic conditions declined during the experiment and the local area was particularly hard hit by redundancies through several employers closing down their operations (history)? Conversely, even if there was some difference between $\mathbf{O_2}$ and $\mathbf{O_1}$, it may be incorrect to conclude that the campaign was effective if history was not controlled, because the experimental effects might have been confounded by history. The longer the time interval between observations, the greater the possibility that history will confound an experiment of this type.[13]

Maturation

Maturation
An extraneous variable attributable to changes in the test units themselves that occur with the passage of time.

Maturation (MA) is similar to history except that it refers to changes in the test units themselves. These changes are not caused by the impact of independent variables or treatments but occur with the passage of time. In an experiment involving people, maturation takes place as people become older, more experienced, tired, bored or uninterested. Tracking and market studies that span several months are vulnerable to maturation, since it is difficult to know how respondents are changing over time.

Maturation effects also extend to test units other than people. For example, consider the case in which the test units are banks. Banks change over time in terms of personnel, physical layout, decoration, and the range of products and services they have to offer.

Testing effects

Testing effects
Effects caused by the process of experimentation.

Testing effects are caused by the process of experimentation. Typically, these are the effects on the experiment of taking a measure on the dependent variable before and after the presentation of the treatment. There are two kinds of testing effects: (1) main testing effect (MT), and (2) interactive testing effect (IT).

Main testing effect
An effect of testing occurring when a prior observation affects a later observation.

The **main testing effect** (MT) occurs when a prior observation affects a later observation. Consider an experiment to measure the effect of advertising on attitudes towards a brand of beer. The respondents are given a pre-treatment questionnaire measuring background information and attitude towards the brand. They are then exposed to the test commercial embedded in a television programme. After viewing the commercial, the respondents again answer a questionnaire measuring, among other things, attitude towards the beer brand.

Suppose that there is no difference between the pre- and post-treatment attitudes. Can we conclude that the commercial was ineffective? An alternative explanation might be that the respondents tried to maintain consistency between their pre- and post-treatment attitudes. As a result of the main testing effect, post-treatment attitudes were influenced more by pre-treatment attitudes than by the treatment itself. The main testing effect may also be reactive, causing the respondents to change their attitudes simply because these attitudes have been measured. The main testing effect compromises the internal validity of the experiment.

Interactive testing effect
An effect in which a prior measurement affects the test unit's response to the independent variable.

In the **interactive testing effect** (IT), a prior measurement affects the test unit's response to the independent variable. Continuing with our beer advertising experiment, when people are asked to indicate their attitudes towards a brand, they become aware of that brand: they are sensitised to that brand and become more likely to pay attention to the test commercial than are people who were not included in the experiment. The measured effects are then not generalisable to the population; therefore, the interactive testing effects influence the experiment's external validity.

Instrumentation

Instrumentation
An extraneous variable involving changes in the measuring instrument, in the observers, or in the scores themselves.

Instrumentation (I) refers to changes in the measuring instrument, in the observers or in the scores themselves. Sometimes measuring instruments are modified during the course of an experiment. In the beer advertising experiment, using a newly designed questionnaire to measure the post-treatment attitudes could lead to variations in the responses obtained. Consider an experiment in which sales at a shoe shop are measured before and after exposure to a promotional offer of a discounted music festival ticket (treatment). A non-experimental price change between $\mathbf{O}_1$ and $\mathbf{O}_2$ results in a change in instrumentation, because sales will be measured using different unit prices. In this case, the treatment effect ($\mathbf{O}_2 - \mathbf{O}_1$) could be attributed to a change in instrumentation.

As shown above, instrumentation effects are likely when interviewers make pre- and post-treatment measurements. The effectiveness of interviewers can be different at different times.

Statistical regression

Statistical regression
An extraneous variable that occurs when test units with extreme scores move closer to the average score during the course of the experiment.

Statistical regression (SR) effects occur when test units with extreme scores move closer to the average score during the course of the experiment. In the beer advertising experiment, suppose that in a pre-test measurement some respondents had either very favourable or very unfavourable attitudes towards the brand. On post-treatment measurement, their attitudes might have moved towards the average. Consumer attitudes change continuously for a wide variety of reasons. Consumers with extreme attitudes have more room for change, so variation may be more likely. This has a confounding effect on the experimental results, because the observed effect (change in attitude) may be attributable to statistical regression rather than to the treatment (test commercial).

Selection bias

Selection bias
An extraneous variable attributable to the improper assignment of test units to treatment conditions.

Selection bias (SB) refers to the improper assignment of test units to treatment conditions. This bias occurs when selection or assignment of test units results in treatment groups that differ on the dependent variable before the exposure to the treatment condition. If test units self-select their own groups or are assigned to groups on the basis of the researchers' judgement, selection bias is possible. For example, consider an experiment in which two different displays (old *static display* and new *audio-visual display*) are assigned to different bank branches. The banks in the two groups may not be equivalent initially. They may vary with respect to an essential characteristic, such as branch size, which is likely to affect the sales of personal loans, regardless of which display was assigned to a bank.

Mortality

Mortality
An extraneous variable attributable to the loss of test units while the experiment is in progress.

Mortality (MO) refers to the loss of test units while the experiment is in progress. This happens for many reasons, such as test units refusing to continue in the experiment. Mortality confounds results because it is difficult to determine whether the lost test units would respond in the same manner to the treatments as those that remain. Consider again the merchandising display experiment. Suppose that during the course of the experiment branch managers in three banks in the new *audio-visual display* drop out because they feel the noise is not conducive to negotiations with certain types of client. The researcher could not determine whether the average sales of the personal loans for the new display would have been higher or lower if these three banks had continued in the experiment.

The various categories of extraneous variables are not mutually exclusive; they can occur jointly and also interact with each other. To illustrate, testing–maturation–mortality refers to a situation in which, because of pre-treatment measurement, the respondents' beliefs and attitudes change over time and there is a differential loss of respondents from the various treatment groups.

Controlling extraneous variables

Confounding variables
Variables used to illustrate that extraneous variables can confound the results by influencing the dependent variable; synonymous with extraneous variables.

Extraneous variables represent alternative explanations of experimental results. They pose a serious threat to the internal and external validity of an experiment. Unless they are controlled, they affect the dependent variable and thus confound the results. For this reason, they are also called **confounding variables**. There are four ways of controlling extraneous variables: randomisation, matching, statistical control and design control.

Randomisation

Randomisation
A method of controlling extraneous variables that involves randomly assigning test units to experimental groups by using random numbers. Treatment conditions are also randomly assigned to experimental groups.

Randomisation refers to the random assignment of test units to experimental groups by using random numbers. Treatment conditions are also randomly assigned to experimental groups. For example, respondents are randomly assigned to one of three experimental groups. One of the three versions of a test commercial, selected at random, is administered to each group. As a result of random assignment, extraneous factors can be represented equally in each treatment condition. Randomisation is the preferred procedure for ensuring the prior equality of experimental groups,[14] but it may not be effective when the sample size is small because it merely produces groups that are equal on average. It is possible, though, to check whether randomisation has been effective by measuring the possible extraneous variables and comparing them across the experimental groups.

Matching

Matching
A method of controlling extraneous variables that involves matching test units on a set of key background variables before assigning them to the treatment conditions.

Matching involves comparing test units on a set of key background variables before assigning them to the treatment conditions. In the display experiment, banks could be matched on the basis of turnover, size, proportion of retail to corporate clients, or location. Then one bank from each matched pair would be assigned to each experimental group.

Matching has two drawbacks. First, test units can be matched on only a few characteristics, so the test units may be similar on the variables selected but unequal on others. Second, if the matched characteristics are irrelevant to the dependent variable, then the matching effort has been futile.[15]

Statistical control

Statistical control
A method of controlling extraneous variables by measuring the extraneous variables and adjusting for their effects through statistical methods.

Statistical control involves measuring the extraneous variables and adjusting for their effects through statistical analysis. This was illustrated in Table 11.2, which examined the relationship (association) between purchase of shares and education, controlling for the effect of income. More advanced statistical procedures, such as analysis of covariance (ANCOVA), are also available. In ANCOVA, the effects of the extraneous variable on the dependent variable are removed by an adjustment of the dependent variable's mean value within each treatment condition. ANCOVA is discussed in more detail in Chapter 19.

Design control

Design control
A method of controlling extraneous variables that involves using specific experimental designs.

Design control involves the use of experiments designed to control specific extraneous variables. The types of controls possible by suitably designing the experiment are illustrated with the following example.

example

Experimenting with new products[16]

Controlled-distribution electronic test markets are used increasingly to conduct experimental research on new products. This method makes it possible to create a design that controls for several extraneous factors. The control can allow for the manipulation of variables that can affect the success of new products. In manipulating variables, it is possible to ensure that a new product:

- obtains the right level of supermarket acceptance and all commodity volume distribution,
- is positioned in the correct aisle in each supermarket,
- receives the right number of facings on the shelf,
- has the correct everyday price,
- never has out-of-stock problems, and
- obtains the planned level of trade promotion, display and price features on the desired time schedule.

By being able to control these variables, a high degree of internal validity can be obtained. ■

A classification of experimental designs

Pre-experimental designs
Designs that do not control for extraneous factors by randomisation.

True experimental designs
Experimental designs distinguished by the fact that the researcher can randomly assign test units to experimental groups and also randomly assign treatments to experimental groups.

Quasi-experimental designs
Designs that apply part of the procedures of true experimentation yet lack full experimental control.

Statistical designs
Designs that allow for the statistical control and analysis of external variables.

Experimental designs may be classified as **pre-experimental**, **true experimental**, **quasi-experimental**, and **statistical designs**: see Figure 11.1.

Pre-experimental designs do not employ randomisation procedures to control for extraneous factors. Examples of these designs include the one-shot case study, the one-group pre-test–post-test design, and the static group. In true experimental designs, the researcher can randomly assign test units to experimental groups and treatments to experimental groups. Included in this category are the pre-test–post-test control group design, the post-test-only control group design, and the Solomon four-group design. Quasi-experimental designs result when the researcher is unable to achieve full manipulation of scheduling or allocation of treatments to test units but can still apply part of the apparatus of the experimentation. Two such designs are time series and multiple time series designs. A statistical design is a series of basic experiments that allows for statistical control and analysis of external variables. Statistical designs are classified

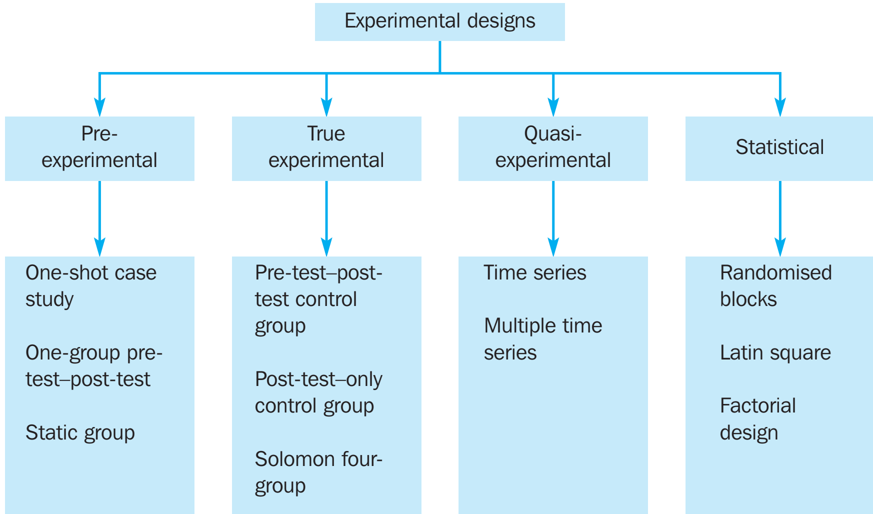

Figure 11.1
A classification of experimental designs

based on their characteristics and use. The important statistical designs include randomised block design, Latin square design and factorial designs.[17]

We begin our discussion with the first type of experimental design: pre-experimental.

Pre-experimental designs

These designs are characterised by an absence of randomisation. Three specific designs are described: the one-shot case study, the one-group pre-test–post-test design, and the static group design.

One-shot case study

One-shot case study
A pre-experimental design in which a single group of test units is exposed to a treatment X, and then a single measurement of the dependent variable is taken.

Also known as the after-only design, the **one-shot case study** may be symbolically represented as

$$\mathbf{X} \quad \mathbf{O_1}$$

A single group of test units is exposed to a treatment **X**, and then a single measurement on the dependent variable is taken ($\mathbf{O_1}$). There is no random assignment of test units. Note that the symbol **R** is not used, because the test units are self-selected or selected arbitrarily by the researcher.

The danger of drawing valid conclusions from experiments of this type can be easily seen. They do not provide a basis of comparing the level of $\mathbf{O_1}$ with what would happen when **X** was absent. In addition, the level of $\mathbf{O_1}$ might be affected by many extraneous variables, including history, maturation, selection and mortality. Lack of control for these extraneous variables undermines the internal validity. For these reasons, the one-shot case study is more appropriate for exploratory than for conclusive research.

One-group pre-test–post-test design

One-group pre-test–post-test design
A pre-experimental design in which a group of test units is measured twice.

The **one-group pre-test–post-test design** may be symbolised as

$$\mathbf{O_1} \quad \mathbf{X} \quad \mathbf{O_2}$$

In this design, a group of test units is measured twice. There is no control group. First a pre-treatment measure is taken ($\mathbf{O_1}$), then the group is exposed to the treatment (**X**). Finally, a post-treatment measure is taken ($\mathbf{O_2}$). The treatment effect is computed as ($O_2 - O_1$) but the validity of this conclusion is questionable since extraneous variables are largely uncontrolled. History, maturation, testing (both main and interactive testing effects), instrumentation, selection, mortality and regression could possibly be present. The following example shows how this design is used.

example

Cinematic performance

It is possible to use the one-group pre-test–post-test design to measure the effectiveness of test commercials. Respondents are recruited to central cinema locations in different test cities. At the central location, respondents are first administered a personal interview to measure, among other things, attitudes towards the brand being portrayed in the commercial (O_1). Then they watch a TV programme containing the test commercial (**X**). After viewing the TV programme, the respondents are again administered a personal interview to measure attitudes towards the same brand (O_2). The effectiveness of the test commercial is measured as ($O_2 - O_1$). ■

Static group design

Static group
A pre-experimental design in which there are two groups: the experimental group (EG), which is exposed to the treatment, and the control group (CG). Measurements on both groups are made only after the treatment, and test units are not assigned at random.

The **static group** is a two-group experimental design. One group, called the experimental group (EG), is exposed to the treatment, and the other, called the control group (CG), is not. Measurements on both groups are made only after the treatment, and test units are not assigned at random. This design may be symbolically described as

$$\begin{array}{lll} \text{EG:} & \mathbf{X} & \mathbf{O_1} \\ \text{CG:} & & \mathbf{O_2} \end{array}$$

The treatment effect would be measured as ($O_1 - O_2$). Notice that this difference could also be attributed to at least two extraneous variables (selection and mortality). Because test units are not randomly assigned, the two groups (EG and CG) may differ before the treatment, and selection bias may be present. There may also be mortality effects, as more test units may withdraw from the experimental group than from the control group. This would be particularly likely to happen if the treatment were unpleasant.

In practice, a control group is sometimes defined as the group that receives the current level of marketing activity, rather than a group that receives no treatment at all. The control group is defined in this way because it is difficult to reduce current marketing activities such as advertising and personal selling to zero. We illustrate the static group in the context of the GlobalCash Project.

example
GlobalCash Project

Static group

A static group comparison to measure the effectiveness of a product placement (a particular character overtly using a branded product or service as part of the story) for a bank, within the context of a feature film, would be conducted as follows. Two groups of respondents would be recruited on the basis of convenience. Only the experimental group would be exposed to the feature film containing the product placement. Then, attitudes towards the bank of both the experimental and control group respondents would be measured. The effectiveness of the product placement would be measured as ($O_1 - O_2$). ■

True experimental designs

The distinguishing feature of true experimental designs, compared with pre-experimental designs, is randomisation. In true experimental designs, the researcher randomly assigns test units to experimental groups and treatments to experimental groups. True experimental designs include the pre-test–post-test control group design, the post-test-only control group design, and the Solomon four-group design.

Pre-test–post-test control group design

Pre-test–post-test control group design
An experimental design in which the experimental group is exposed to the treatment but the control group is not. Pre-test and post-test measures are taken on both groups.

In the **pre-test–post-test control group design**, test units are randomly assigned to either the experimental or the control group and a pre-treatment measure is taken on each group. Then the treatment is applied to the experimental group, and a post-treatment measure is taken from both groups. This design is symbolised as:

EG:	**R**	O_1	**X**	O_2
CG:	**R**	O_3		O_4

The treatment effect (TE) is measured as

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

This design controls for most extraneous variables. Selection bias is eliminated by randomisation. The other extraneous effects are controlled as follows:

$$O_2 - O_1 = TE + H + MA + MT + IT + I + SR + MO$$
$$O_4 - O_3 = H + MA + MT + I + SR + MO$$
$$= EV \text{ (extraneous variables)}$$

where the symbols for the extraneous variables are as defined previously. The experimental result is obtained by

$$(O_2 - O_1) - (O_4 - O_3) = TE + IT$$

Interactive testing effect is not controlled, because of the effect of the pre-test measurement on the reaction of units in the experimental group to the treatment.

example
GlobalCash Project

Pre-test–post-test control group

In the context of measuring the effectiveness of a product placement in a feature film for a bank, a pre-test–post-test control group design would be implemented as follows. A sample of respondents would be selected at random. Half of these would be randomly assigned to the experimental group, and the other half would form the control group. Respondents in both groups would be administered a questionnaire to obtain a pre-test measurement on attitudes towards the bank. Only the respondents in the experimental group would be exposed to the feature film containing the product placement. Then, a questionnaire would be administered to respondents in both groups to obtain post-test measures on attitudes towards the bank. ■

As this example shows, the pre-test–post-test control group design involves two groups and two measurements on each group. A simpler design is the post-test-only control group design.

Post-test-only control group design

Post-test-only control group design
Experimental design in which the experimental group is exposed to the treatment but the control group is not and no pre-test measure is taken.

The **post-test-only control group design** does not involve any pre-measurement. It may be symbolised as

EG:	**R**	**X**	O_1
CG:	**R**		O_2

The treatment effect is obtained by

$$O_1 - O_2 = TE$$

This design is fairly simple to implement. Because there is no pre-measurement, the testing effects are eliminated, but this design is sensitive to selection bias and mortality. It is assumed that the two groups are similar in terms of pre-treatment measures on the dependent variable because of the random assignment of test units to groups.

Because there is no pre-treatment measurement, this assumption cannot be checked. This design is also sensitive to mortality. It is difficult to determine whether those in the experimental group who discontinue the experiment are similar to their counterparts in the control group. Yet another limitation is that this design does not allow the researcher to examine changes in individual test units.

It is possible to control for selection bias and mortality through carefully designed experimental procedures. Examination of individual cases is often not of interest. On the other hand, this design possesses significant advantages in terms of time, cost and sample size requirements. It involves only two groups and only one measurement per group. Because of its simplicity, the post-test-only control group design is probably the most popular design in marketing research. Note that, except for pre-measurement, the implementation of this design is very similar to that of the pre-test–post-test control group design.

Solomon four-group design

Solomon four-group design
An experimental design that explicitly controls for interactive testing effects, in addition to controlling for all the other extraneous variables.

In this case, the researcher is not concerned with examining the changes in the attitudes of individual respondents. When this information is desired, the **Solomon four-group design** should be considered. The Solomon four-group design overcomes the limitations of the pre-test–post-test control group and post-test-only control group designs in that it explicitly controls for interactive testing effect, in addition to controlling for all the other extraneous variables. However, this design has practical limitations: it is expensive and time-consuming to implement. Hence, it is not considered further.[18]

In all true experimental designs, the researcher exercises a high degree of control. In particular, the researcher can control when the measurements are taken, on whom they are taken, and the scheduling of the treatments. Moreover, the researcher can randomly select the test units and randomly expose test units to the treatments. In some instances, the researcher cannot exercise this kind of control; then quasi-experimental designs should be considered.

Quasi-experimental designs

A quasi-experimental design results under the following conditions. First, the researcher can control when measurements are taken and on whom they are taken. Second, the researcher lacks control over the scheduling of the treatments and also is unable to expose test units to the treatments randomly.[19] Quasi-experimental designs are useful because they can be used in cases when true experimentation cannot be used, and because they are quicker and less expensive. Because full experimental control is lacking, the researcher must consider the specific variables that are not controlled. Popular forms of quasi-experimental designs are time series and multiple time series designs.

Time series design

Time series design
A quasi-experimental design that involves periodic measurements of the dependent variable for a group of test units. Then the treatment is administered by the researcher or occurs naturally. After the treatment, periodic measurements are continued to determine the treatment effect.

The **time series design** involves a series of periodic measurements on the dependent variable for a group of test units. The treatment is then administered by the researcher or occurs naturally. After the treatment, periodic measurements are continued to determine the treatment effect. A time-series experiment may be symbolised as:

$$O_1\ O_2\ O_3\ O_4\ O_5\ O_6\ O_7\ O_8\ O_9\ O_{10}$$

This is a quasi-experiment, because there is no randomisation of test units to treatments, and the timing of treatment presentation, as well as which test units are exposed to the treatment, may not be within the researcher's control (hence there being no specific **X** symbolised above).

Taking a series of measurements before and after the treatment provides at least partial control for several extraneous variables. Maturation is at least partially controlled, because it would not affect $\mathbf{O_5}$ and $\mathbf{O_6}$ alone but would also influence other observations. By similar reasoning, main testing effect and statistical regression are controlled as well. If the test units are selected randomly or by matching, selection bias can be reduced. Mortality may pose a problem, but it can be largely controlled by paying a premium or offering other incentives to respondents.

The major weakness of the time series design is the failure to control history. Another limitation is that the experiment may be affected by the interactive testing effect because multiple measurements are being made on the test units. Nevertheless, time series designs are useful, as illustrated by this case. The effectiveness of a test commercial (**X**) may be examined by broadcasting the commercial a predetermined number of times and examining the data from a pre-existing test panel. Although the marketer can control the scheduling of the test commercial, it is uncertain when or whether the panel members are exposed to it. The panel members' purchases before, during and after the campaign are examined to determine whether the test commercial has a short-term effect, a long-term effect or no effect.

Multiple time series design

Multiple time series design
A time series design that includes another group of test units to serve as a control group.

The **multiple time series design** is similar to the time series design except that another group of test units is added to serve as a control group. Symbolically, this design may be described as

$$\begin{array}{lllllllllllll} \text{EG:} & \mathbf{O_1} & \mathbf{O_2} & \mathbf{O_3} & \mathbf{O_4} & \mathbf{O_5} & \mathbf{X} & \mathbf{O_6} & \mathbf{O_7} & \mathbf{O_8} & \mathbf{O_9} & \mathbf{O_{10}} \\ \text{CG:} & \mathbf{O_{11}} & \mathbf{O_{12}} & \mathbf{O_{13}} & \mathbf{O_{14}} & \mathbf{O_{15}} & & \mathbf{O_{16}} & \mathbf{O_{17}} & \mathbf{O_{18}} & \mathbf{O_{19}} & \mathbf{O_{20}} \end{array}$$

If the control group is carefully selected, this design can be an improvement over the simple time series experiment. The improvement lies in the ability to test the treatment effect twice: against the pre-treatment measurements in the experimental group and against the control group. To use the multiple time series design to assess the effectiveness of a commercial, the test panel example would be modified as follows. The test commercial would be shown in only a few of the test cities. Panel members in these cities would make up the experimental group. Panel members in cities where the commercial was not shown would constitute the control group.

Another application of multiple time series design is illustrated in the following example.

example

Splitting commercials shows their strength[20]

A multiple time series design was used to examine the build-up effect of increased advertising. The data were obtained from Burke Marketing Services from a split-cable TV advertising field experiment. In the split-cable system, one group of households was assigned to the experimental panel and an equivalent group was assigned to the control panel. The two groups were matched on demographic variables. Data were collected for 76 weeks. Both panels received the same level of advertising for the first 52 weeks for the brand in question. For the next 24 weeks, the experimental panel was exposed to twice as much advertising as the control panel. The results indicated that the build-up effect of advertising was immediate with a duration of the order of the purchase cycle. Information of this type can be useful in selecting advertising timing patterns (allocating a set of advertising exposures over a specified period to obtain maximum impact).

Table 11.3 Potential sources of invalidity of experimental designs

Source of invalidity								
	Internal variables							External variables
Design	*History*	*Maturation*	*Testing*	*Instrumentation*	*Regression*	*Selection*	*Mortality*	*Interaction of testing and X*
Pre-experimental designs								
One-shot case study X O	–	–				–	–	
One-group pre-test–post-test design O X O	–	–	–	–	?			–
Static group comparison X O O	+	?	+	+	+	–	–	
True experimental designs								
Pre-test–post-test control group R O X O R O O	+	+	+	+	+	+	+	–
Post-test only control group design R X O R O	+	+	+	+	+	+	+	+
Quasi-experimental designs								
Time series O O O X O O O	–	+	+	?	+	+	+	–
Multiple time series O O O X O O O O O O O O O	+	+	+	+	+	+	–	

Note: A minus sign indicates a definite weakness, a plus sign indicates that the factor is controlled, a question mark denotes a possible source of concern, and a blank means that the factor is not relevant.

In concluding our discussion of pre-experimental, true experimental and quasi-experimental designs, we summarise in Table 11.3 the potential sources of invalidity that may affect each of these designs. In this table, a minus sign indicates a definite weakness, a plus sign indicates that the factor is controlled, a question mark denotes a possible source of concern, and a blank means that the factor is not relevant. It should be remembered that potential sources of invalidity are not the same as actual errors.

Statistical designs

Statistical designs consist of a series of basic experiments that allow for statistical control and analysis of external variables. In other words, several basic experiments are conducted simultaneously. Thus, statistical designs are influenced by the same sources of invalidity that affect the basic designs being used. Statistical designs offer the following advantages:

1 The effects of more than one independent variable can be measured.
2 Specific extraneous variables can be statistically controlled.
3 Economical designs can be formulated when each test unit is measured more than once.

The most common statistical designs are the randomised block design, the Latin square design and the factorial design.

Randomised block design

Randomised block design
A statistical design in which the test units are blocked on the basis of an external variable to ensure that the various experimental and control groups are matched closely on that variable.

A **randomised block design** is useful when there is only one major external variable – such as sales, store size, or income of the respondent – that might influence the dependent variable. The test units are blocked or grouped on the basis of the external variable. The researcher must be able to identify and measure the blocking variable. By blocking, the researcher ensures that the various experimental and control groups are matched closely on the external variable.

example
GlobalCash Project

Randomised block design

Let us extend the example examining the effectiveness of a product placement in a feature film for a bank. The purpose of this experiment would be to measure the impact of environmental concern in a film character related to the bank. Suppose that a pan-European bank like ABN-AMRO were to sponsor a film that included shots using their buildings, logos and examples of the way they run their business. They would naturally be concerned that the image portrayed in the film enhanced the corporate image that they wish to project.

To test this, three test film clips, A, B and C, show respectively a character in the film with no environmental concern, some environmental concern, and high environmental concern. Which of these would be the most effective? Management feels that the respondents' evaluation of the product placement will be influenced by the extent of their usage of a bank. So bank usage is identified as the blocking variable and the randomly selected respondents are

Character A from the test film clips, displaying 'no environmental concern' by sneaking up to and scaring pigeons.

classified into four blocks (heavy, medium, light, or non-users of the bank). Respondents from each block are randomly assigned to the treatment groups (test film clips A, B, and C). The results reveal that the some environmental concern commercial (B) was most effective overall (see Table 11.4).

Table 11.4 An example of a randomised block design

		Treatment groups		
Block number	Bank usage	Film A	Film B	Film C
1	*High*			
2	*Medium*			
3	*Light*			
4	*None*			

As this example illustrates, in most marketing research situations, external variables such as sales, bank size, bank type, bank location and characteristics of the respondent can influence the dependent variable. Therefore, randomised block designs are generally more useful than completely random designs. Their main limitation is that the researcher can control for only one external variable. When more than one variable must be controlled, the researcher must use Latin square or factorial designs.

Latin square design

Latin square design
A statistical design that allows for the statistical control of two non-interacting external variables in addition to the manipulation of the independent variable.

A **Latin square design** allows the researcher to control statistically two non-interacting external variables as well as to manipulate the independent variable. Each external or blocking variable is divided into an equal number of blocks or levels. The independent variable is also divided into the same number of levels. A Latin square is conceptualised as a table (see Table 11.5), with the rows and columns representing the blocks in the two external variables. The levels of the independent variable are then assigned to the cells in the table. The assignment rule is that each level of the independent variable should appear only once in each row and each column, as shown in Table 11.5.

Table 11.5 An example of a Latin square design

	Interest in increasing electronic automation		
Bank usage	*High*	*Medium*	*Low*
High	B	A	C
Medium	C	B	A
Light and None	A	C	B

Note: A, B and C denote the three test commercials, which have respectively no environmental concern, some concern and high concern.

example
GlobalCash Project

Latin square design

To illustrate the Latin square design, suppose that in the previous example, in addition to controlling for bank usage, the researcher also wanted to control for interest in increasing the electronic automation of cash transactions (defined as high, medium or low). To implement a Latin square design, bank usage would also have to be blocked at three rather than four levels, (e.g. by combining the low and non-users into a single block). Assignments of the three test film clips could then be made as shown in Table 11.5. Note that each film clip – A, B or C – appears once, and only once, in each row and each column.

Although Latin square designs are popular in marketing research, they are not without limitations. They require an equal number of rows, columns and treatment levels, which is sometimes problematic. Note that, in the above example, the low users and non-patrons had to be combined to satisfy this requirement. In addition, only two external variables can be controlled simultaneously. Latin squares do not allow the researcher to examine interactions of the external variables with each other or with the independent variable. To examine interactions, factorial designs should be used.

Factorial design

Factorial design
A statistical experimental design used to measure the effects of two or more independent variables at various levels and to allow for interactions between variables.

A **factorial design** is used to measure the effects of two or more independent variables at various levels. Unlike the randomised block design and the Latin square, factorial designs allow for interactions between variables.[21] An interaction is said to take place when the simultaneous effect of two or more variables is different from the sum of their separate effects. For example, an individual's favourite drink might be coffee and her favourite temperature level might be cold, but this individual might not prefer cold coffee, leading to an interaction.

A factorial design may also be conceptualised as a table. In a two-factor design, each level of one variable represents a row and each level of another variable represents a column. Multidimensional tables can be used for three or more factors. Factorial designs involve a cell for every possible combination of treatment variables. Suppose that in the previous example, in addition to examining the effect of environmental concern, the researcher was also interested in simultaneously examining the effect of the amount of information about the bank that came over in the film clip. Further, the amount of bank information was also varied at three levels (high, medium and low). As shown in Table 11.6, this would require $3 \times 3 = 9$ cells. The respondents would be randomly selected and randomly assigned to the nine cells. Respondents in each cell would receive a specific treatment combination. For example, respondents in the upper left corner cell would view a film clip that had no environmental concern and low bank information. The results revealed a significant interaction between the two factors or variables. Respondents with a low amount of bank information preferred the high environmental concern film clip (C). Those with a high amount of bank information, however, preferred the no environmental concern film clip (A). Notice that, although Table 11.6 may appear somewhat similar to Table 11.4, the random assignment of respondents and data analysis are very different for the randomised block design and the factorial design.[22]

Table 11.6 An example of a factorial design

	Amount of environmental concern		
Amount of bank information	*No concern*	*Some concern*	*High concern*
Low			
Medium			
High			

Another example of a factorial design follows.

example

Price and information are for the dogs[23]

Burke Marketing Research conducted an experiment prior to the launch of a new dog food. They wished to determine the effect of price and competitive brand information on purchase intentions. A two-factor design was used. Price was manipulated to have four levels: one

Sally celebrates being chosen to take part in the experiment prior to the launch of a new dog food.

discount, two parity (or mid-range prices), and one premium. Competitive brand information was varied at two levels: whether or not information on competitive brands was provided. Approximately 240 respondents were randomly assigned to one of eight (4 × 2) treatment conditions. Respondents were asked to indicate their purchase intentions for the new product on a five-point scale. The results indicated that neither price nor competitive brand information had a significant effect on purchase intentions. ■

The main disadvantage of a factorial design is that the number of treatment combinations increases multiplicatively with an increase in the number of variables or levels. In the Burke Marketing Research example, if the price had been manipulated at six levels and competitive brand information at three levels (no information, partial information and full information), the number of cells would have jumped from 8 to 18. All the treatment combinations are required if all the main effects and interactions are to be measured. If the researcher is interested in only a few of the interactions or main effects, fractional factorial designs may be used. As their name implies, these designs consist of only a fraction or portion of the corresponding full factorial design.

Laboratory versus field experiments

Field environment
An experimental location set in actual market conditions.

Laboratory environment
An artificial setting for experimentation in which the researcher constructs the desired conditions.

Experiments may be conducted in a laboratory or field environment. A laboratory environment is an artificial one that the researcher constructs with the desired conditions specific to the experiment. The term **field environment** is synonymous with actual market conditions. Our experiment to measure the effectiveness of a film clip could be conducted in a **laboratory environment** by showing the film in a test cinema. The same experiment could also be conducted in a field environment by running the full test film (rather than clips) in conventional cinemas. The differences between the two environments are summarised in Table 11.7.

Table 11.7 Laboratory versus field experiments

Factor	*Laboratory*	*Field*
Environment	Artificial	Realistic
Control	High	Low
Reactive error	High	Low
Demand artefacts	High	Low
Internal validity	High	Low
External validity	Low	High
Time	Short	Long
Number of units	Small	Large
Ease of implementation	High	Low
Cost	Low	High

Laboratory experiments have the following advantages over field experiments:

- The laboratory environment offers a high degree of control because it isolates the experiment in a carefully monitored environment. Therefore, the effects of history can be minimised.
- A laboratory experiment also tends to produce the same results if repeated with similar subjects, leading to high internal validity.
- Laboratory experiments tend to use a small number of test units, last for a shorter time, be more restricted geographically, and are easier to conduct than field experiments. Hence, they are generally less expensive as well.

Compared with field experiments, laboratory experiments suffer from some main disadvantages:

- The artificiality of the environment may cause reactive error in that the respondents react to the situation itself rather than to the independent variable.[24]
- The environment may cause **demand artefacts**, a phenomenon in which the respondents attempt to guess the purpose of the experiment and respond accordingly. For example, while viewing the film clip, the respondents may recall pre-treatment questions about the brand and guess that the commercial is trying to change their attitudes towards the brand.[25]
- Finally, laboratory experiments are likely to have lower external validity than field experiments. Because a laboratory experiment is conducted in an artificial environment, the ability to generalise the results to the real world may be diminished.

Demand artefacts
Responses given because the respondents attempt to guess the purpose of the experiment and respond accordingly.

It has been argued that artificiality or lack of realism in a laboratory experiment need not lead to lower external validity. One must be aware of the aspects of the laboratory experiment that differ from the situation to which generalisations are to be made. External validity will be reduced only if these aspects interface with the independent variables explicitly manipulated in the experiment, as is often the case in applied marketing research. Another consideration, however, is that laboratory experiments allow for more complex designs than field experiments. Hence, the researcher can control for more factors or variables in the laboratory setting, which increases external validity.[26]

The researcher must consider all these factors when deciding whether to conduct laboratory or field experiments.[27] Field experiments are less common in marketing research than laboratory experiments, although laboratory and field experiments play complementary roles.[28]

Characteristics and limitations of experimental designs

In Chapter 3, we discussed three types of research designs: exploratory, descriptive and causal. Of these, it may be argued that causal designs are the most appropriate for inferring and measuring cause-and-effect relationships (though not the only way – as may be argued by the adherents to grounded research approaches, introduced in Chapter 6). Although descriptive survey data are often used to provide evidence of 'causal' relationships, these studies do not meet all the conditions required for causality. For example, it is difficult in descriptive studies to establish the prior equivalence of the respondent groups with respect to both the independent and dependent variables. On the other hand, an experiment can establish this equivalence by random assignment of test units to groups. In descriptive research, it is also difficult to establish time order of occurrence of variables. In an experiment, however, the researcher controls the timing of the measurements and the introduction of the treatment. Finally, descriptive research offers little control over other possible causal factors.

We do not wish to undermine the importance of descriptive research designs in marketing research. As mentioned in Chapter 3, descriptive research constitutes the most popular research design in marketing research, and we do not want to imply that it should never be used to examine causal relationships. Indeed, some authors have suggested procedures for drawing causal inferences from descriptive (non-experimental) data.[29] Rather, our intent is to alert the reader to the limitations of descriptive research for examining causal relationships. Likewise, we also want to make the reader aware of the limitations of experimentation.[30]

Experimentation is an important research design that gives the ability to infer causal relationships. However, it has limitations of time, cost and administration of an experiment.

Time

Experiments can be time-consuming, particularly if the researcher is interested in measuring the long-term effects of the treatment, such as the effectiveness of an advertising campaign. Experiments should last long enough so that the post-treatment measurements include most or all of the effects of the independent variables.

Cost

Experiments are often expensive. The requirements of experimental group, control group and multiple measurements significantly add to the cost of research.

Administration

Experiments can be difficult to administer. It may be impossible in measuring human activity to control for the effects of the extraneous variables, particularly in a field environment. Field experiments often interfere with a company's ongoing operations, and obtaining cooperation from the retailers, wholesalers and others involved may be difficult. Finally, competitors may deliberately contaminate the results of a field experiment. These limitations have given rise to the use of grounded theory approaches, especially in developing an understanding of consumer behaviour that is impossible to encapsulate through experiments.

Experimental design application: test marketing

Test marketing
An application of a controlled experiment done in limited, but carefully selected, test markets. It involves a replication of the planned national marketing programme for a product in test markets.

Test markets
A carefully selected part of the marketplace particularly suitable for test marketing.

Test marketing, also called market testing, is an application of a controlled experiment conducted in limited but carefully selected parts of the marketplace called **test markets**. It involves a replication of a planned national marketing programme in test markets. Often, the marketing mix variables (independent variables) are varied in test marketing and the sales (dependent variable) are monitored so that an appropriate national marketing strategy can be identified. The two major objectives of test marketing are (1) to determine market acceptance of the product, and (2) to test alternative levels of marketing mix variables. Test-marketing procedures may be classified as standard test markets, controlled and mini-market tests, and simulated test marketing.

Standard test market

Standard test market
A test market in which the product is sold through regular distribution channels. For example, no special considerations are given to products simply because they are being test-marketed.

In a **standard test market**, test markets are selected and the product is sold through regular distribution channels. Typically, the company's own sales force is responsible for distributing the product. Sales personnel stock the shelves, restock, and take inventory at regular intervals. One or more combinations of marketing mix variables (product, price, distribution and promotional levels) are employed.

Designing a standard test market involves deciding what criteria are to be used for selecting test markets, how many test markets to use, and the duration of the test. Test markets must be carefully selected. In general, the more test markets that can be used, the better. If resources are limited, at least two test markets should be used for each programme variation to be tested. Where external validity is important, however, at least four test markets should be used. The criteria for selection of test markets may be summarised as:[31]

1. Large enough to produce meaningful projections. They should contain at least 2% of the potential target population.
2. Representative demographically.
3. Representative with respect to product consumption behaviour.
4. Representative with respect to media usage.
5. Representative with respect to competition.
6. Relatively isolated in terms of media and physical distribution.
7. Having normal historical development in the product class.
8. Having marketing research and auditing services available.
9. Not overtested.

The duration of the test depends on the repurchase cycle for the product, the probability of competitive response, cost considerations, the initial consumer response, and company philosophy. The test should last long enough for repurchase activity to be observed. This indicates the long-term impact of the product. If competitive reaction to the test is anticipated, the duration should be short. The cost of the test is also an important factor. The longer a test, the more it costs, and at some point the value of additional information is outweighed by its costs. Recent evidence suggests that tests of new brands should run for at least 10 months. An empirical analysis found that the final test market share was reached in 10 months 85% of the time and in 12 months 95% of the time. Test marketing can be very beneficial to a product's successful introduction, as the following example demonstrates.

example

Bass joins exclusive Czech beer club[32]

Bass has acquired 34% of Staropramen, a Prague brewer. It launched the Czech beer in six-month test markets in Manchester and Liverpool in bottles, and on draft and in bottles in London. The introduction was backed with a comprehensive promotional package designed to encourage consumer trial and future purchase. This included sampling nights, point-of-sale material and glassware. ■

A standard test market, such as the Bass example, constitutes a one-shot case study. In addition to the problems associated with this design, test marketing faces two unique problems. First, competitors often take actions such as increasing their promotional efforts to contaminate the test marketing programme. When Procter & Gamble test-marketed a hand-and-body lotion, the market leader, Cheeseborough Ponds, started a competitive buy-one-get-one-free promotion for its flagship brand, Vaseline Intensive Care lotion. This encouraged consumers to stock up on Vaseline Intensive Care lotion and, as a result, the Procter & Gamble product did poorly in the test market. Procter & Gamble still launched the line nationally. Ponds again countered with the same promotional strategy. Vaseline Intensive Care settled with a market share of 22% while Procter & Gamble achieved but 4%.[33] Another problem is that, while a firm's test marketing is in progress, competitors have an opportunity to beat it to the national market[34].

Sometimes it is not feasible to implement a standard test market using the company's personnel. Instead, the company must seek help from an outside supplier, in which case the controlled test market may be an attractive option.

Controlled test market

Controlled test market
A test-marketing programme conducted by an outside research company in field experimentation. The research company guarantees distribution of the product in retail outlets that represent a predetermined percentage of the market.

In a **controlled test market**, the entire test-marketing programme is conducted by an outside research company. The research company guarantees distribution of the product in retail outlets that represent a predetermined percentage of the market. It handles warehousing and field sales operations, such as stocking shelves, selling and stock control. The controlled test market includes both mini-market (or forced distribution) tests and the smaller controlled store panels. This service is provided by a number of research firms, including A.C. Nielsen.

Simulated test market

Simulated test market
A quasi-test market in which respondents are preselected; they are then interviewed and observed on their purchases and attitudes towards the product.

Also called a laboratory test or test market simulation, a **simulated test market** yields mathematical estimates of market share based on initial reaction of consumers to a new product. The procedure works as follows. Typically, respondents are intercepted in busy locations, such as shopping centres, and pre-screened for product usage. The selected individuals are exposed to the proposed new product concept and given an opportunity to buy the new product in a real-life or laboratory environment. Those who purchase the new product are interviewed about their evaluation of the product and repeat purchase intentions. The trial and repeat-purchase estimates so generated are combined with data on proposed promotion and distribution levels to project a share of the market.[35]

Simulated test markets can be conducted in 16 weeks or fewer. The information they generate is confidential and the competition cannot obtain it. They are also relatively inexpensive. Simulated test markets can cost around 10% of a standard test market.

Determining a test-marketing strategy

The first decision to be made is whether or not to test market the proposed new product, or whatever element of the marketing programme that is under consideration. As shown in Figure 11.2, this decision must take into account the competitive environ-

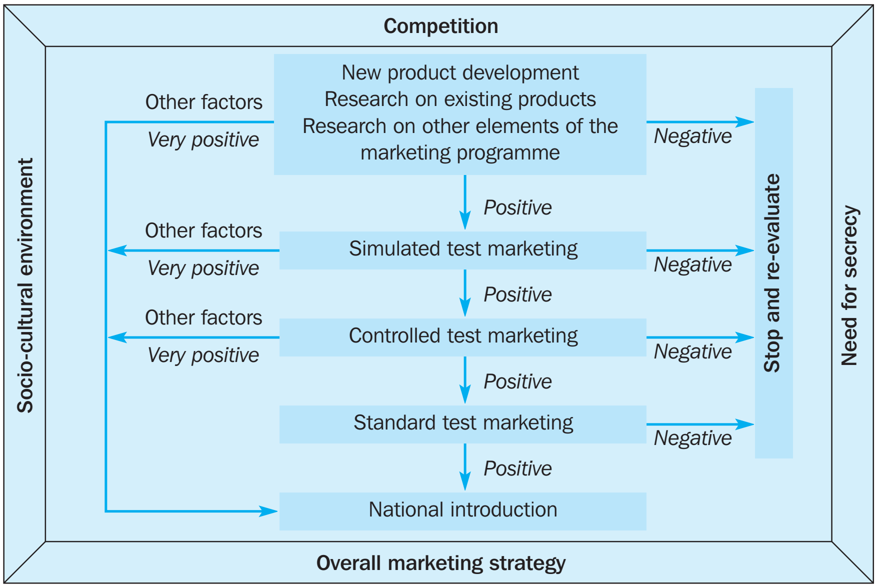

Figure 11.2
Selecting a test-marketing strategy

ment; the socio-cultural environment, particularly consumer preferences and past behaviours; the need to keep the firm's marketing efforts secret; and the overall marketing strategy of the firm. If the marketing research already undertaken to develop the new product provides compelling positive evidence, or if factors such as pre-empting competitive moves dominate, the new product may well be introduced nationally without test marketing. If the decision is to conduct test marketing, however, simulated test marketing may be conducted first, followed by controlled test marketing, then standard test marketing, and, if the results are positive, national introduction. Of course, very positive results at any stage may directly lead to national introduction, circumventing subsequent testing.

International marketing research

If field experiments are difficult to conduct in developed Western economies, the challenge they pose is greatly increased in the international arena. In many countries, the marketing, economic, structural, information and technological environment is not developed to the extent that it is in Europe and the United States. For example, in many countries, TV stations are owned and operated by a government that may place severe restrictions on television advertising. This makes field experiments that manipulate advertising levels extremely difficult. Consider, for example, M & M/Mars, which has set up massive manufacturing facilities in Russia and advertises its sweets on television. Yet, the sales potential has not been realised. Is Mars advertising too much, too little or just enough? Although the answer could be determined by conducting a field experiment that manipulated the level of advertising, such causal research is not feasible given the tight control by the Russian government on television stations. Likewise, the lack of major supermarkets in the Baltic states makes it difficult for Procter & Gamble to conduct field experiments to determine the effect of in-store promotions on the sales of its detergents.

In some countries in Asia, Africa and South America, a majority of the population live in small towns and villages. Yet basic infrastructure such as roads, transportation and warehouse facilities are lacking, making it difficult to achieve desired levels of distribution. Even when experiments are designed in such countries, it is difficult to control for the time order of occurrence of variables and the absence of other possible causal factors, two of the necessary conditions for causality. Because the researcher has little control over the environment, control of extraneous variables is particularly problematic. Furthermore, it may not be possible to address this problem by adopting the most appropriate experimental design, as environmental constraints may make that design infeasible.

Thus, the internal and external validity of field experiments conducted overseas is generally lower than in Europe and the United States. Although pointing to the difficulties of conducting field experiments in other countries, we do not wish to imply that such causal research cannot or should not be conducted. On the contrary, as the following example indicates, creatively designed field experiments can result in rich findings.

example

What you hear is what you get[36]

PepsiCo's strategy to fight arch-rival Coca-Cola in France was through increased spending on advertisements. Part of Coca-Cola's campaign was to use singers and celebrities such as Rod Stewart, Tina Turner, Gloria Estefan and M.C. Hammer in their commercials as well as publicity tours. Marketing research, however, revealed that overplaying American celebrities was detrimental in the French market. Pepsi thought that this was probably a weakness Coke had because Europeans considered Coca-Cola's marketing effort as 'too American' in Europe. Pepsi, therefore, decided to use taste as a competitive tool. The key was to highlight the product superiority, although not directly, as comparative advertising was prohibited in France. They came up with music as a means of communicating. How did this work?

Research showed that attitude towards the brand was influenced by attitude towards the ad, especially in low-involvement products such as soft drinks. Sweet and melodious music played for Pepsi could transfer good feelings from the music to the Pepsi brand. Similarly, repugnant and undesirable music played for Coke could also transfer from the music to the Coke brand. This mechanism is called classical conditioning. To test these hypotheses, a two-factor experiment was designed. The two factors would be the type of music and the brand preferred, each varied at two levels as shown. A test commercial for each experimental condition was run for a month, with each commercial being played in a different city. At the end of the campaign, central location interviews were used to examine the effect on the brand.

		Brand type	
		Pepsi	*Coke*
Music type	*Good*		
	Bad		

The results of a similar experiment indicated positive effects for 'good' music and negative effects for 'bad' music. Pepsi designed its advertising based on these findings. Subsequently, retail sales in France increased, although it remained in second position, after Coke. ■

Ethics in marketing research

As was explained in Chapter 10, it is often believed that, if respondents are aware of the purpose of a research project, they may give biased responses. In these situations, a deliberate attempt is made by the researcher to disguise the purpose of the research. This is often necessary with experimentation, where disguise is needed to produce valid results. Take, for example, a project conducted to determine the effectiveness of television commercials for a breakfast cereal. The respondents are recruited and brought to a central facility. They are told that they will be watching a television programme on nutrition and then will be asked some questions. Interspersed in the programme is a test commercial for the breakfast cereal as well as commercials for some other products (filler commercials). After viewing the programme and the commercials, the respondents are given a questionnaire to complete. The questionnaire obtains evaluations on the programme content, the test commercial and some of the filler commercials. Note that the evaluations of the programme content and the filler commercials are not of interest but are obtained to reinforce the nature of the disguise. If the respondents knew that the true purpose was to determine the effectiveness of the test commercial, their responses might be biased. Disguising the purpose of the research, however, should not lead to deception.

Although this seems like a paradox, one solution is to disclose the possible existence of deception before the start of the experiment and allow the participants the right to redress at the conclusion of the experiment. The following four items should be conveyed: (1) inform respondents that in an experiment of this nature a disguise of the purpose is often required for valid results; (2) inform them of the general nature of the experiment and what they will be asked to do; (3) make sure they know that they can leave the experiment at any time; and (4) inform them that the study will be fully explained after the data have been gathered and at that time they may request that their information be withdrawn.

Debriefing
After a disguised experiment, informing test subjects what the experiment was about and how the experimental manipulations were performed.

The procedure outlined in item (4) is called **debriefing**. It could be argued that disclosure in this way would also bias results. There is evidence, however, indicating that data collected from subjects informed of the possibility of deception and those not informed are similar.[37] Debriefing can alleviate the stress caused by the experiment and make the experiment a learning experience for the respondents. However, if not handled carefully, debriefing itself can be unsettling to subjects. In the breakfast cereal example above, respondents may find it disheartening that they spent their time evaluating a cereal commercial. The researcher should anticipate and address this issue in the debriefing session.

One further ethical concern in experimentation involves using the appropriate experimental design to control errors caused by extraneous variables. It is the responsibility of the researcher to use the most applicable experimental design for the problem. As the following example illustrates, determining the most appropriate experimental design for the problem requires not only an initial evaluation but also continuous monitoring.

example

Correcting errors early: a stitch in time saves nine[38]

A marketing research firm specialising in advertising research examined the effectiveness of a television commercial for Nike athletic shoes. A one-group pre-test–post-test design was used. Attitudes held by the respondents towards Nike athletic shoes were obtained prior to being exposed to a sports programme on TV and several commercials, including the one for

Nike. Attitudes were again measured after viewing the programme and the commercials. Initial evaluation based on a small sample found the one-group pre-test–post-test design adopted in this study to be susceptible to demand artefacts: respondents attempt to guess the purpose of the experiment and respond accordingly. Because time and financial constraints make redesigning the study difficult at best, the research continued with correction. Continuing a research project after knowing errors were made in the early stages is not ethical behaviour. Experimental design problems should be disclosed immediately to the client. Decisions whether to redesign or accept the flaws should be made jointly. ■

Internet and computer applications

The Internet can be a useful vehicle for conducting causal research. Different experimental treatments can be displayed on different Websites. Respondents can then be recruited to visit these sites and respond to a questionnaire that obtains information on the dependent and extraneous variables. Thus, the Internet can provide a mechanism for controlled experimentation, although in a laboratory type of environment.

An example of testing the effectiveness of advertisements can be used to illustrate the use of the Internet in causal research. Different advertisements can be posted on different Websites. Matched or randomly selected respondents can be recruited to visit these sites, with one group visiting only one site. If any pre-treatment measures have to be obtained, respondents can answer a questionnaire posted on the site. Then they are exposed to a particular advertisement on that site. After viewing the advertisement, the respondents answer additional questions providing post-treatment measures. Control groups can also be implemented in a similar way. Thus, all types of experimental designs that we have considered can be implemented in this manner.

To complement the Internet, microcomputers and mainframe software can be used in the design and analysis of experiments. For example, the statistical analysis package Minitab can be used to design experiments – see www.minitab.com for details and to download a demo. Minitab is similar in use to SPSS (www.spss.com) or SAS (www.sas.com). Minitab includes functions and documentation specifically for industrial-quality control work in which factorial designs are encountered. For example, researchers investigating restaurant atmosphere might want to examine some of the interactions of independent variables. The dependent variable in this experiment could be the respondent's rating of the restaurant as a setting for a romantic meal. Three factors would be included in this $2 \times 2 \times 2$ study. Assuming two lighting levels (i.e. low or medium), two sound types (piped music or live music), and two olfactory stimuli (i.e. spicy smells or sweet/confectionery smells), the best combinations of restaurant atmospherics can be examined.

The marketing research company A. C. Nielsen are major proponents of test-marketing designs. Visit their Website on www.acnielsen.com, click on 'test marketing' and review their approaches under the headings of 'Live/In-store New Product Testing, 'Live/In-Store Existing Product Testing', and 'Simulated Product Testing – BASES'.

Summary

The scientific notion of causality implies that we can never prove that *X* causes *Y*. At best, we can only infer that *X* is one of the causes of *Y* in that it makes the occurrence of *Y* probable. Three conditions must be satisfied before causal inferences can be made: (1) concomitant variation, which implies that *X* and *Y* must vary together in a hypothesised way; (2) time order of occurrence of variables, which implies that *X* must precede *Y*; and (3) elimination of other possible causal factors, which implies that competing explanations must be ruled out. Experiments provide the most convincing evidence of all three conditions. An experiment is formed when one or more independent variables are manipulated or controlled by the researcher and their effect on one or more dependent variables is measured.

In designing an experiment, it is important to consider internal and external validity. Internal validity refers to whether the manipulation of the independent variables actually caused the effects on the dependent variables. External validity refers to the generalisability of experimental results. For the experiment to be valid, the researcher must control the threats imposed by extraneous variables, such as history, maturation, testing (main and interactive testing effects), instrumentation, statistical regression, selection bias and mortality. There are four ways of controlling extraneous variables: randomisation, matching, statistical control and design control.

Experimental designs may be classified as pre-experimental, true experimental, quasi-experimental and statistical designs. An experiment may be conducted in a laboratory environment or under actual market conditions in a real-life setting. Only causal designs encompassing experimentation are appropriate for inferring cause-and-effect relationships.

Although experiments have limitations in terms of time, cost and administration, they are becoming increasingly popular in marketing. Test marketing is an important application of experimental design.

The internal and external validity of field experiments conducted in developing nations is generally lower than in the developed Western economies. The level of development in many countries is lower and the researcher lacks control over many of the marketing variables. The ethical issues involved in conducting causal research include disguising the purpose of the experiment. Debriefing can be used to address some of these issues.

Questions

1. What are the requirements for inferring a causal relationship between two variables?
2. Differentiate between internal and external validity.
3. List any five extraneous variables and give an example to show how each can reduce internal validity.
4. Describe the various methods for controlling extraneous sources of variation.
5. What is the key characteristic that distinguishes true experimental designs from pre-experimental designs?
6. List the steps involved in implementing the post-test-only control group design. Describe the design symbolically.
7. What is a time series experiment? When is it used?

8 How is a multiple time series design different from a basic time series design?

9 What advantages do statistical designs have over basic designs?

10 What are the limitations of the Latin square design?

11 Compare the characteristics of laboratory and field experimentation.

12 Should descriptive research be used for investigating causal relationships? Why or why not?

13 What is test marketing? What are the three types of test marketing?

14 What is the main difference between a standard test market and a controlled test market?

15 Describe how simulated test marketing works.

Notes

1 Dwek, R., 'Prediction of success', *Marketing* (POP and Field Marketing Supplement) (17 April 1997), XII–XIII; 'POP radio test airs the ads in-store', *Marketing News* (24 October 1986), 16.

2 Bausall, R.B., *Conducting Meaningful Experiments* (Thousand Oaks, CA: Sage, 1994).

3 For several references on the use of experiments in marketing, see Gardner, D.M. and Belk, R.W., *A Basic Bibliography on Experimental Design in Marketing* (Chicago, IL: American Marketing Association, 1980).

4 Boruch, R.F., *Randomized Experiments for Planning and Evaluation* (Thousand Oaks, CA: Sage, 1994).

5 Witthaus, M., 'POP stars', *Marketing Week* 20(16) (17 July 1997), 37–41.

6 Wyner, G.A., 'Experimental design', *Marketing Research: A Magazine of Management and Applications* 9(3) (Fall 1997), 39–41; Brown, S.R. and Melamed, L.E., *Experimental Design and Analysis* (Newbury Park, CA: Sage, 1990).

7 To see a study employing experimental designs: Spence, M.T. and Brucks, M., 'The moderating effect of problem characteristics on experts' and novices' judgments', *Journal of Marketing Research* 34 (May 1997), 233–47.

8 Love, A.A., 'Companies want to cut coupons, but consumers demand bargains', *Marketing News* 31(10) (12 May 1997), 15; Shoemaker, R.W. and Tibrewala, V., 'Relating coupon redemption rates to past purchasing of the brand', *Journal of Advertising Research* 25 (October–November 1985), 40–7.

9 Banks, S., *Experimentation in Marketing* (New York: McGraw-Hill, 1965), 168–79.

10 In addition to internal and external validity, there also exist construct and statistical conclusion validity. Construct validity addresses the question of what construct, or characteristic, is in fact being measured and is discussed in Chapter 12 on measurement and scaling. Statistical conclusion validity addresses the extent and statistical significance of the covariation that exists in the data and is discussed in the chapters on data analysis. See also: Cook, T.D. and Campbell, D.T., *Quasi-Experimentation* (Chicago, IL: Rand McNally, 1979), 52–3; and Campbell, D.T. and Stanley, J.C., *Experimental and Quasi Experimental Designs for Research* (Chicago, IL: Rand McNally, 1966).

11 Bordia, P., 'Face-to-face computer-mediated communication: a synthesis of the experimental literature', *Journal of Business Communication* 34(1) (January 1997), 99–120; Bowen, D.M., 'Work group research: past strategies and future opportunities', *IEEE Transactions on Engineering Management* 42(1) (February 1995), 30–8; Lynch, J.G., Jr, 'On the external validity of experiments in consumer research', *Journal of Consumer Research* 9 (December 1982), 225–44.

12 Argyris, C., 'Actionable knowledge: design causality in the service of consequential theory', *Journal of Applied Behavioural Science* 32(4) (December 1966), 390-406; Lynch, J.G., Jr, 'The role of external validity in theoretical research', Calder, B.J., Phillips, L.W. and Tybout, A., 'Beyond external validity', and McGrath, J.E and Brinberg, D., 'External validity and the research process', *Journal of Consumer Research* (June 1983), 109–11, 112–14 and 115–24.

13 Hirsch, A.R., 'Effects of ambient odors on slot machine usage in a Las Vegas casino', *Psychology and Marketing* 12(7) (October 1995), 585–94.

14 Durier, C., Monod, H. and Bruetschy, A., 'Design and analysis of factorial sensory experiments with carry-over effects', *Food Quality and Preference* 8(2) (March 1997), 141–9; Nelson, L.S., 'Notes on the use of randomization in experimentation', *Journal of Quality Technology* 28(1) (January 1996), 123–6.

15 Selart, M., 'Structure compatibility and restructuring in judgement and choice', *Organisation Behaviour and Human Decision Processes* 65(2) (February 1996), 106–16; Barker Bausell, R., *Conducting Meaningful Experiments* (Thousand Oaks, CA: Sage, 1994).

16 Spethmann, B., 'Choosing a test market', *Brandweek* 36(19) (8 May 1995), 42–3; Tarshis, A.M., 'Natural sell-in avoids pitfalls of controlled tests', *Marketing News* (24 October 1986), 14.

17 Other experimental designs are also available. See Gunter, B., 'Fundamental issues in experimental design', *Quality Progress* 29(6) (June 1996), 105–13; Winer, R.S., 'Analysis of advertising experiments', *Journal of Advertising Research* (June 1980), 25–31.

18 For an application of the Solomon four-group design, see Mizerski, R.W., Allison, N.K. and Calvert, S., 'A controlled field study of corrective advertising using multiple exposures and a commercial medium', *Journal of Marketing Research* 17 (August 1980), 341–8.

19 Moorman, C., 'A quasi-experiment to assess the consumer and informational determinants of nutrition information-processing activities – the case of the Nutrition Labeling and Education Act', *Journal of Public Policy and Marketing* 15(1) (Spring 1996), 28–44.

20 Lodish, L.M., Abraham, M.M., Livelsberger, J. and Lubetkin, B., 'A summary of fifty-five in-market experimental estimates of the long-term effects of TV advertising', *Marketing Science* (Summer 1995), G133–40; Krishnamurthi, L., Narayan, J. and Raj, S.P., 'Intervention analysis of a field experiment to assess the build-up effect of advertising', *Journal of Marketing Research* 23 (November 1986), 337–45.

21 For applications of factorial designs, see Bryce, W.J., Day, R. and Olney, T.J., 'Commitment approach to motivating community recycling: New Zealand curbside trial', *Journal of Consumer Affairs*, 31(1) (Summer 1997), 27–52; Leclerc, F., Schmitt, B.H. and Dube, L., 'Foreign branding and its effects on product perceptions and attitudes', *Journal of Marketing Research* 31 (May 1994), 263–70.

22 Jones, G.E. and Kavanagh, M.J., 'An experimental examination of the effects of individual and situational factors on unethical behavioural intentions in the workplace', *Journal of Business Ethics* 15(5) (May 1996), 511–23; Bemmaor, A.C. and Mouchoux, D., 'Measuring the short-term effect of in-store promotion and retail advertising on brand sales: a factorial experiment', *Journal of Marketing Research* 28 (May 1991), 202–14.

23 Dickson, J.R. and Wilby, C.P., 'Concept testing with and without product trial', *Journal of Product Innovation Management* 14(2) (March 1997), 117–25; Miller, J.B., Bruvold, N.T. and Kernan, J.B., 'Does competitive-set information affect the results of concept tests?', *Journal of Advertising Research* (April–May 1987), 16–23.

24 Bone, P.F., 'Word-of-mouth effects on short-term and long-term product judgments', *Journal of Business Research* 32(3) (March 1995), 213–23; Barnes, J.H. Jr and Seymour, D.T., 'Experimenter bias: task, tools, and time', *Journal of the Academy of Marketing Science* (Winter 1980), 1–11.

25 Perrien, J., 'Repositioning demand artifacts in consumer research', *Advances in Consumer Research* 24, (1997) 267–71; Shrimp, T.A., Hyatt, E.M. and Snyder, D.J., 'A critical appraisal of demand artifacts in consumer research', *Journal of Consumer Research* 18(3) (December 1991), 272–83.

26 Monden, Y., Akter, M. and Kubo, N., 'Target costing performance based on alternative participation and evaluation method: a laboratory experiment', *Managerial and Decision Economics* 18(2) (March 1997), 113–29.

27 Farris, P.W. and Reibstein, D.J., 'Overcontrol in advertising experiments', *Journal of Advertising Research* (June–July 1984), 37–42.

28 Alston, R.M. and Nowell, C., 'Implementing the voluntary contribution game: a field experiment', *Journal of Economic Behaviour and Organisation* 31(3) (December 1996), 357–68; and Houston, M.J. and Rothschild, M.L., 'Policy-related experiments on information provision: a normative model and explication', *Journal of Marketing Research* (November 1980), 432–49.

29 Spanos, A., 'On theory testing in econometrics: modeling with nonexperimental data', *Journal of Econometrics* 67(1) (May 1995), 189–226; Blalock, H.M. Jr, *Causal Inferences in Non-experimental Research* (Chapel Hill, NC: University of North Carolina Press, 1964).

30 In some situations, surveys and experiments can complement each other and may both be used. For example, the results obtained in laboratory experiments may be further examined in a field survey. See Johnston, W.J. and Kim, K., 'Performance, attribution, and expectancy linkages in personal selling', *Journal of Marketing* 58 (October 1994), 68–81.

31 Romeo, P., 'Testing, testing', *Restaurant Business* 97(2) (15 January 1998), 12. Reprinted with permission from *Marketing News* published by the American Marketing Association, Chicago, 1 March 1985, p. 15.

32 'Bass joins exclusive Czech beer club', *Grocer* 216 (7174) (22 October 1994), 34.

33 Mehegan, S., 'Vaseline ups ante via anti-bacterial', *Brandweek* 38(21) (26 May 1997), 1, 6.

34 'Why new products are bypassing the market test', *Management Today* (October 1995), 12.

35 Hayes, D.J., Shogren, J.F., Fox, J.A. and Kliebenstein, B., 'Test marketing new food products using a multitrial nonhypothetical experimental auction', *Psychology and Marketing* 13(4) (July 1996), 365–79.

36 Crumley, B., 'French cola wars', *Advertising Age* (17 December 1990), 22.

37 Schmitt, B.H., 'Contextual priming of visual information in advertisements', *Psychology and Marketing* 11(1) (January/February 1994), 1–14.

38 Peterson, B., 'Ethics revisited', *Marketing Research: A Magazine of Management and Applications* 8(4) (Winter 1996), 47–8.