Chapter 2
DEFINING THE RESEARCH DESIGN AND CONTROLLING RESEARCH ERRORS

A research design specifies the methods and procedures for acquiring the information needed to structure and solve the research problem. The overall operational design for a research project stipulates what information is to be collected, from what sources, and by what procedures. A good research design ensures that the information obtained is relevant to the research problem, and that it is collected by objective and economical procedures. A research design might be described as a series of advance decisions that, taken together, form a master plan or model for conducting a research investigation.

Research designs vary depending on the type of study. Generally designs are associated with three types of studies, those that focus on providing exploratory research, descriptive research and causal research. Each will be described in turn.

Exploratory Studies

The major purposes of exploratory studies are for the identification of problems, the precise formulation of problems (including the identification of relevant variables), and the formulation of new alternative courses of action.

An exploratory study is often the first in a series of projects. That is, an exploratory study is often used as an introductory phase of a larger study, and its results are used to bring focus to the larger study and to develop specific techniques that will be used. Thus flexibility is a key to designing and conducting exploratory studies.

We can distinguish three separate tasks that are usually included in exploratory studies and that are typically conducted in the sequence listed:

- A search of secondary information sources
- Interviews with persons knowledgeable about the subject area
- The examination of analogous situations

Search Secondary Sources

Secondary sources of information are the "literature" on the subject. It is the rare research problem for which there is no relevant information to be found by a relatively quick and inexpensive search of the literature. Secondary information sources are not limited to external sources. Searches should also be made of company records.

Interview Knowledgeable Persons

Having searched secondary sources, it is usually desirable to talk with persons who are well informed in the area being investigated, including company executives, experts, consumers and mavens, and users outside the organization

A widely used technique in exploratory research is the focus group. In focus group interviews, a group of knowledgeable people participate in a joint interview that does not use a structured question-and-answer methodology. The group, usually consisting of 8 to 12 people (but may have as few as 5 or as many as 20), is purposely selected to include individuals who have a common background, or similar buying or use experience, as related to the problem being researched. The interviewer or moderator of the focus group session works with the client to develop a general discussion outline that typically includes such topics as usage experience, problems with use, and how decisions are made. The objective is to foster involvement and interaction among the group members during the interview that will lead to spontaneous discussion and the disclosure of attitudes, opinions, and information on present or prospective buying and use behavior.

Focus groups are used primarily to identify and define problems, provide background information, and generate hypotheses. Focus groups typically do not provide solutions for problems. Areas of application include detecting trends in lifestyles, examining new product or service concepts, generating ideas for improving established products or services, developing creative concepts for advertising, and determining effective means of marketing products or services.

If the sole purpose is to create ideas, then individual interviews may be a better alternative than focus groups. Limited research on this issue conducted more than 20 years ago suggests that the number and quality of ideas generated may be greater from such interviews (Fern, 1982).

More specific uses of focus groups include:

1. Identifying and understanding consumer language relating to the product category in question. What terms do they use? What do they mean?

2. Identifying the range of consumer concerns. How much variability is there among consumers' perception of the product, and in the considerations leading them to accept or reject the product?

3. Identifying the complexity of consumer concerns. Do a few simple attitudes govern consumer reaction toward the product, or is the structure complex, involving many contingencies?

4. Identifying specific methodological or logistical problems that are likely to affect either the cost of the subsequent research, or one's ability to generate meaningful, actionable findings.

An example of focus group usage might be to determine the reasons for the decline in a product's overall rating, as reported in a syndicated research report.

Examine Analogous Situations

It is also logical that a researcher will want to examine analogous situations to determine what else can be learned about the nature of the problem and its variables. Analogous situations include case histories and simulations. More discussion of the use of focus groups is given in Chapter 4.

Descriptive Studies

Much research is concerned with describing market characteristics or functions. For example, a market potential study may describe the number, distribution, and socioeconomic characteristics of potential customers of a product. A market-share study finds the share of the market received by both the company and its major competitors. A sales analysis study describes sales by territory, type of account, size or model of product, and the like. In marketing, descriptive studies are also made in the following areas:

- Product research: identification and comparison of functional features and specifications of competitive products
- Promotion research: description of the demographic characteristics of the audience being reached by the current advertising program
- Distribution research: determining the number and location of retailers handling the company's products that are supplied by wholesalers versus those supplied by the company's distribution centers
- Pricing research: identifying competitors' prices by geographic area

These examples of descriptive research cover only a few of the possibilities. Descriptive designs, often called observational designs by some researchers, provide information on groups and phenomena that already exist; no new groups are created (Fink, 2003).

One example of a descriptive study is one conducted by a school-employees credit union in order to gain information useful to provide better service to its members. Management knew very little about the members, other than that they were school employees, family members of employees, or former employees. In addition, the credit union knew very little about member awareness and use of, and attitudes toward individual services available to them. Consequently, investigators undertook a study to answer the following research questions:

1. What are the demographic and socioeconomic characteristics of primary members?
2. How extensively are existing services being used, and what are members' attitudes toward such services?
3. What is the degree of interest in specific new services?

Although associations can be used only to make inferences, and not establish a causal relationship, they are often useful for predictive purposes. It is not always necessary to understand causal relations in order to make accurate predictive statements. Descriptive information often provides a sound basis for the solution of marketing problems, even though it does not explain the nature of the relationship involved. The basic principle involved is to find desirable behavior correlates that are measurable when the predictive statement is made.

Causal Studies

Although descriptive information is often useful for predictive purposes, where possible we would like to know the causes of what we are predicting—the "reasons why." Further, we would like to know the relationships of these causal factors to the effects that we are predicting. If we understand the causes of the effects we want to predict, we invariably improve our ability both to predict and to control these effects.

Bases for Inferring Causal Relationships

There are three types of evidence that can be used for drawing inferences about causal relationships:

1. Associative variation
2. Sequence of events
3. Absence of other possible causal factors

In addition, the cause and effect have to be related. That is, there must be logical implication (or theoretical justification) to imply the specific causal relation.

Associative Variation

Associative variation, or "concomitant variation," as it is often termed, is a measure of the extent to which occurrences of two variables are associated. Two types of associative variation may be distinguished:

1. Association by presence: A measure of the extent to which the presence of one variable is associated with the presence of the other

2. Association by change: A measure of the extent to which a change in the level of one variable is associated with a change in the level of the other.

It has been argued that two other conditions may also exist, particularly for continuous variables: (a) the presence of one variable is associated with a change in the level of other; and (b) a change in the level of one variable is associated with the presence of the other (Feldman, 1975).

Sequence of Events

A second characteristic of a causal relationship is the requirement that the causal factor occur first; the cause must precede the result. In order for salesperson retraining to result in increased sales, the retraining must have taken place prior to the sales increase.

Absence of Other Possible Causal Factors

A final basis for inferring causation is the absence of any other possible causes other than the one(s) being investigated. If it could be demonstrated, for example, that no other factors present could have caused the sales increase in the third quarter, we could then logically conclude that the salesperson training must have been responsible.

Obviously, in an after-the-fact examination of an uncontrolled result such as an increase in detergent sales, it is impossible to clearly rule out all causal factors other than salesperson retraining. One could never be completely sure that there were no competitor-, customer-, or company-initiated causal factors that would account for the sales increase.

Conclusions Concerning Types of Evidence

No single type of evidence, or even the combination of all three types considered, can ever conclusively demonstrate that a causal relationship exists. Other unknown factors may exist. However, we can obtain evidence that makes it highly reasonable to conclude that a particular relationship exists. Exhibit 2.1 shows certain questions that are necessary to answer.

EXHIBIT 2.1 Issues in Determining Causation

Several questions arise when determining whether a variable *X* has causal power over another variable, *Y:*

1. What is the source of causality—does *X* cause *Y,* or does *Y* cause *X?*
2. What is the direction of causality—does *X* positively influence *Y,* or is the relationship negative?
3. Is *X* a necessary and sufficient cause—or necessary, but not sufficient cause—of *Y?* Is *X*'s causation deterministic or probabilistic?
4. Which value of the believed cause exerts a causal influence—its presence or absence?
5. Are the causes and effects the states themselves or changes in the states? Is the relationship static or dynamic?

In the end, the necessary conditions for causality to exist are a physical basis for causality, a cause that temporally precedes the effect (even for associative variation), and a logical reason to imply the specific causal relation being examined. (Monroe and Petroshius, n.d.).

SOURCES OF MARKETING INFORMATION

There are five major sources for obtaining marketing information. In this section we briefly describe each as an introduction to subsequent chapters that describe some of these sources in more depth.

- Secondary sources
- Respondents
- Natural experiments
- Controlled experiments
- Simulation

Secondary Sources of Information

Secondary information is information that has been collected by persons or agencies for purposes other than the solution of the problem at hand.

> If a furniture manufacturer, for example, needs information on the potential market for furniture in the Middle Atlantic States, many government and industry sources of secondary information are available.
>
> The federal government collects and publishes information on the numbers of families, family formation, income, and the number and sales volume of retail stores, all by geographic area. It also publishes special reports on the furniture industry. Many state and local governments collect similar information for their respective areas.
>
> The trade associations in the furniture field collect and publish an extensive amount of information about the industry. Trade journals are also a valuable source of secondary information, as are special studies done by other advertising media.

Private research firms collect specialized marketing information on a continuing basis and sell it to companies. These so-called syndicated services, particularly those for packaged consumer goods, are becoming more sophisticated as they are increasingly becoming based on scanner data. Technology advancements are having a measurable impact on the availability of secondary data.

Information from Respondents

A second major source of information is obtained from respondents. Asking questions and observing behavior are primary means of obtaining information whenever people's actions are being investigated or predicted.

The term respondent literally means "one who responds or answers."
Both verbal and behavioral responses should be considered.

In this book we shall consider both the information obtained from asking people questions, and that provided by observing behavior (or the results of past behavior) to comprise information from respondents.

Information from Natural and Controlled Experiments

As described earlier, three types of evidence provide the bases for drawing inferences about causal relationships. While both natural and controlled experimental designs are capable of providing associative variation and sequence of events, only controlled experiments can provide reasonably conclusive evidence concerning the third type of evidence, the absence of other possible producers.

A natural experiment is one in which the investigator intervenes only to the extent required for measurement. That is, there is no manipulation of an assumed causal variable. The investigator merely looks at what has happened. As such, the natural experiment is a form of ex post facto research. In this type of study, the researcher approaches data collection as if a controlled experimental design were used. The variable of interest has occurred in a natural setting, and the researcher looks for respondents who have been exposed to it and also, if a control group is desired, respondents who have not been exposed.

Measurements can then be made on a dependent variable of interest. For example, if the impact of a television commercial on attitudes were desired, the investigator would conduct a survey of people after the commercial was shown. Those who saw the commercial would constitute the experimental group, and those who did not see it would be a type of control group. Differences in attitudes could be compared as a crude measure of impact. Unfortunately, one can never be sure whether the obtained relationship is causal or non-causal, since the attitudes may be affected by the presence of other variables. For a brief discussion of natural experiments, see Anderson (1971).

In controlled experiments, investigator intervention is required beyond that needed for measurement purposes. Specifically, two kinds of intervention are required:

1. Manipulation of at least one assumed causal variable
2. Random assignment of subjects to experimental and control groups

The researcher conducts the experiment by assigning the subjects to an experimental group where the causal variable is manipulated, or to a control group where no manipulation of the causal variable occurs. The researcher measures the dependent variable in both situations and then tests for differences between the groups. Differences between the groups, if present, are attributed to the manipulation variable.

Field experiments are increasingly being completed using online survey instruments. For example, researchers often use the advanced branching logic, randomization, question block presentation, question timing, and java scripting capabilities of Qualtrics.com to conduct time and cost effective field experiments.

Simulation

The expense and time involved in the personal interviews often associated with field experimentation may preclude it as a source of information for a particular operational situation. In such cases it may be desirable to construct a model of the operational situation and to experiment with it instead of venturing into a real-world situation. The manipulation of such models is called simulation.

Simulation can be defined as a set of techniques for manipulating a model of some real-world process to find numerical solutions that represent the real process being modeled. Models that are environmentally rich (that is, that contain many complex interactions and nonlinear relationships among the variables, probabilistic components, time dependencies, etc.) are usually too difficult to solve by standard analytical methods such as calculus or other mathematical programming techniques. Rather, the analyst views a simulation model as a limited imitation of the process or system under study and attempts to run the system on a computer to see what would happen if a particular policy were put into effect.

Simulations may be used for research, instruction, decision-making, or some combination of these applications. During the past 50 or more years, simulations have been developed for such marketing decision- making applications as marketing systems, marketing-mix elements (new-product, price advertising, and sales-force decisions), and interviewing costs in marketing surveys.

TYPES OF ERRORS AFFECTING RESEARCH DESIGNS

The marketing research process (and research design) involves the management of error. Potential errors can arise at any point from problem formulation through report preparation, and rarely will a research project be error-free. Consequently, the research designer must adopt a strategy for managing and minimizing this error. As we shall see in the next section of this chapter, there are alternative strategies that can be followed.

The objective underlying any research project is to provide information that is as accurate as possible. Maximizing accuracy requires that total study errors be minimized. Total study error has two components—sampling error and non-sampling error—and can be expressed as follows:

Total error = Sampling error + Non-sampling error

Total error is usually measured as total error variance, also known as the mean-squared error (Assael & Keon, 1982):

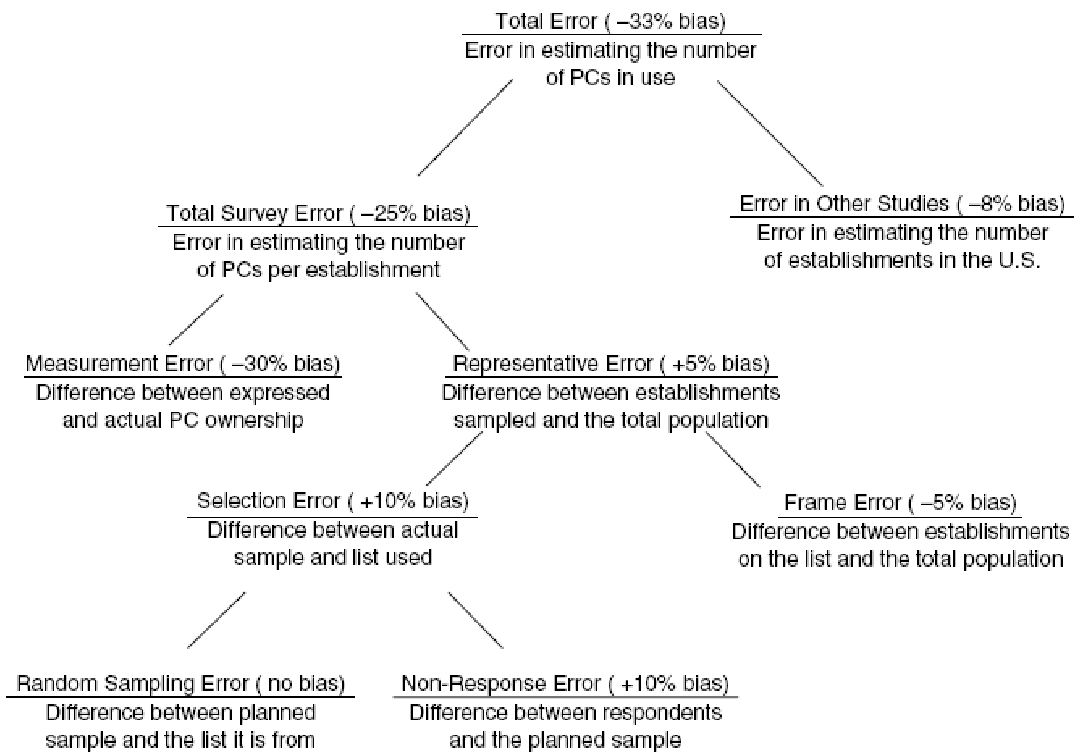
$$(\text{Total error})^2 = (\text{Sampling error})^2 + (\text{Non-sampling error})^2$$

Sampling error refers to the variable error resulting from the chance specification of population from elements according to the sampling plan. Since this introduces random variability into the precision with which a sample statistic is calculated, it is often called random sampling error. Exhibit 2.2 gives an illustration of how total error is assessed.

EXHIBIT 2.2 How Errors Add Up

It is important to know all the sources of error that contribute to inaccuracy, and to assess the impact of each. As an example, consider the figure below, which shows components of error in a study designed to estimate the size of the personal computer market (Lilien, Brown, & Searls, 1991).

When estimating the market, adjustments are made for each source of error. The components are then combined mathematically to create the total error. For purposes of simplicity, total error is shown here as the sum of the component errors. In actuality, total error would be smaller, as it is usually based on the square roots of summed squares of component errors. Assessing the individual components of total error is highly judgmental and subjective, but it is worth the effort.



SOURCE: Reprinted with permission from "How Errors Add Up," by Lilien, G., Brown, R., & Searls, K. in *Marketing News, 33,* January 7, 1991. Published by the American Marketing Association.

Non-sampling error consists of all other errors associated with a research project. Such errors are diverse in nature and are often thought of as resulting in some sort of bias, which implies systematic error. Bias can be defined simply as the difference between the true value of that which is being measured and the researchers' estimate of the true value. However, there can be a random component of non-sampling error. For example, misrecording a response during data collection would represent a random error, whereas using a loaded question would be a systematic error. Non-sampling errors have both nonresponse and response based origins.

To a large extent these major error components are inversely related. Increasing the sample size to reduce sampling error can increase non-sampling error in that, for example, there are more instances where such things as recording errors can occur, and the impact of biased (i.e., nonobjective) questions and other systematic errors will be greater. Thus, this inverse relationship lies at the heart of our concern for total error.

Ideally, efforts should be made to minimize each component. Considering time and cost limitations this can rarely be done. The researcher must make a decision that involves a tradeoff between sampling and non-sampling errors. Unfortunately, very little is known empirically about the relative size of the two error components, although there is some evidence that non-sampling error tends to be the larger of the two. In a study comparing several research designs and data collection methods, Assael and Keon (1982) concluded that non-sampling error far outweighs random sampling error in contributing to total survey error.

As an introduction, Exhibit 2.3 briefly defines eight major types of errors that can influence research results.

EXHIBIT 2.3 Types of Errors in the Research Process

Different types of errors can influence research results:

- **Population specification:** noncorrespondence of the required population to the population selected by the researcher
- **Sampling:** noncorrespondence of the sample selected by probability means and the representative sample sought by the researcher
- **Selection:** noncorrespondence of the sample selected by nonprobability means and the sought representative sample
- **Frame:** noncorrespondence of the sought sample to the required sample
- **Nonresponse:** noncorrespondence of the achieved (or obtained) sample to the selected sample
- **Surrogate information:** noncorrespondence of the information being sought by the researcher and that required to solve the problem
- **Measurement:** noncorrespondence of the information obtained by the measurement process and the information sought by the researcher
- **Experimental:** noncorrespondence of the true (or actual) impact of, and the impact attributed to, the independent variable(s)

Population Specification Error

This type of error occurs when the researcher selects an inappropriate population or universe from which to obtain data.

Examples: Cessna Aircraft conducts an online survey to learn what features should be added to a proposed corporate jet. They consider conducting a survey of purchasing agents from major corporations presently owning such aircraft. However, they believe that that this would be an inappropriate research universe; since pilots are most likely play a key role in the purchase decision.

Similarly, packaged goods manufacturers often conduct surveys of housewives, because they are easier to contact, and it is assumed they decide what is to be purchased and also do the actual purchasing. In this situation there often is population specification error. The husband may purchase a significant share of the packaged goods, and have significant direct and indirect influence over what is bought.

Sampling Error

Sampling error occurs when a probability sampling method is used to select a sample, but the resulting sample is not representative of the population concern.

> **Example**: Suppose that we collected a random sample of 500 people from the general adult population and upon analysis found it to be composed only of people aged 35 to 55. This sample would not be representative of the general adult population. Sampling error is affected by the homogeneity of the population being studied and sampled from and by the size of the sample.

In general, the more homogeneous the population (meaning smaller variance on any given characteristic of interest), the smaller the sampling error; as sample size increases, sampling error decreases. If a census were conducted (i.e., all elements of the population were included) there would be no sampling error.

Selection Error

Selection error is the sampling error for a sample selected by a nonprobability method.

> **Example**: Interviewers conducting a mall intercept study have a natural tendency to select those respondents who are the most accessible and agreeable whenever there is latitude to do so. Such samples often comprise friends and associates who bear some degree of resemblance in characteristics to those of the desired population.

Selection error often reflects people, who are most easily reached, better dressed, and have better kept homes or more pleasant personalities. Samples of these types rarely are representative of the desired population.

Frame Error

A sampling frame is the source for sampling that represents all the members of the population. It is usually a listing of the prospective respondents to be sampled.

> **Example**: Consider the sample frame for a shopper intercept study at a shopping mall. This sample frame includes all shoppers in the mall during the period of data collection. A commonly used frame for consumer research is the telephone directory. This frame introduces error because many elements of the population are not included in the directory (unlisted phone numbers, new arrivals), some elements are listed more than once, and nonpopulation elements are also included (businesses, people who have left the area).

A perfect frame identifies each member of the population once, but only once, and does not include members not in the population of interest.

Nonresponse Error

Nonresponse error can exist when an obtained sample differs from the original selected sample. There are two ways in which nonresponse can occur: (a) noncontact (the inability to contact all members of the sample); and (b) refusal (nonresponse to some or all items on the measurement instrument). Errors arise in virtually every survey from the inability to reach respondents.

Example: In telephone surveys, some respondents are inaccessible because they are not at home (NAH) for the initial call or call-backs. Others have moved or are away from home for the period of the survey. Not-at-home respondents are typically younger with no small children, and have a much higher proportion of working wives than households with someone at home. People who have moved or are away for the survey period have a higher geographic mobility than the average of the population. Thus, most surveys can anticipate errors from non-contact of respondents.

Refusals may be by item or for the entire interview. Income, religion, sex, and politics are topics that may elicit item refusals. Some respondents refuse to participate at all because of time requirements, health issues, past experiences in which an "interviewer" turned out to be a telemarketer, or other reasons. Refusals can also be specific to the method of data collection, as in nonresponse to a mail and email questionnaires or using caller ID to screen and avoid telephone surveys. Nonresponse to mail and email questionnaires sometimes runs as high as 90 percent of the initial mailing, even after several successive mailings.

The amount of effort involved in data collection is another possible way to affect nonresponse error. However, little research has been done to examine the impact of effort.

Example: In a national telephone survey, a so-called five-day "standard" survey was compared to a "rigorous" survey conducted over an eight-week period (Keeter, Miller, Kohut, Groves, & Presser, 2000). Response rates were significantly different; the rigorous survey generated about two-thirds greater response. But the two surveys produced similar results. Most of the statistically significant differences were for demographic items. Very few differences were found on substantive variables.

Nonresponse is also a potential problem in business-to-business and within organization research situations. Although specific respondents are individuals, organizations are not, as they are differentiated and hierarchical. These characteristics may affect organizational response to survey requests.

Tomaskovic-Devey, Leiter, and Thompson (1994) in a study of organizational response, stated the likelihood that an organizational respondent will respond is a function of three characteristics of the respondent:

1. Authority to respond: The degree to which a designated respondent has the formal or informal authority to respond to a survey request
2. Capacity to respond: Organizational practices and the division of labor and information affect the assembly of relevant knowledge to reply adequately
3. Motive to respond: Both individual and organizational motivations to provide information (or not provide information) about the organization.

Surrogate Information Error

In many research situations, it is necessary to obtain information that acts as a surrogate for that which is required. The necessity to accept substitute information arises from either the inability or unwillingness of respondents to provide the information requested.

Decision-oriented behavioral research is always concerned with the prediction of behavior. This limits most marketing research projects to using proxy information, since one cannot observe future behavior. Typically, researchers obtain one or more kinds of surrogate information believed to be useful in predicting behavior.

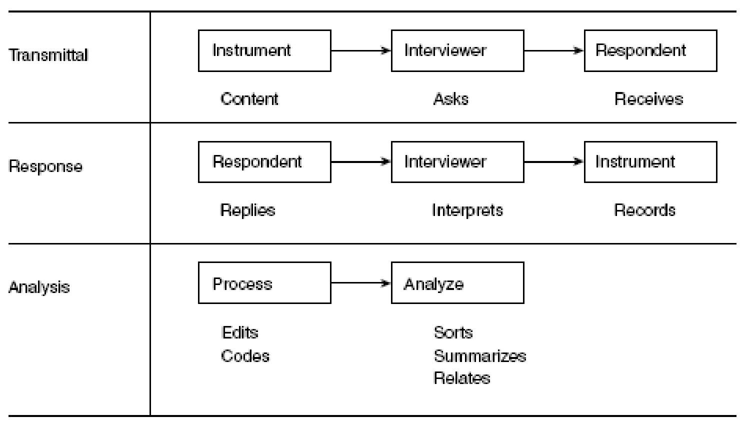
> **Examples**: One may obtain information on past behavior because it is believed that there is sufficient stability in the underlying behavior pattern to give it reasonably high predictive validity. One may ask about intended behavior as a means of prediction. Or one may obtain information about attitudes, level of knowledge, or socioeconomic characteristics of the respondent in the belief that, individually or collectively, they have a high degree of association with future behavior.

Since the type of information required is identified during the problem-formulation stage of the research process, minimizing this error requires an accurate problem definition.

Measurement Error

Measurement error is generated by the measurement process itself, and represents the difference between the information generated and the information wanted by the researcher. Such error can potentially arise at any stage of the measurement process, from the development of an instrument through the analysis of the findings. To illustrate, Figure 2.1 depicts the stages at which errors in eliciting information may arise when interviewing respondents for a survey.

FIGURE 2.1 Potential Sources of Measurement Error in a Survey



In the transmittal stage, errors may be due to the faulty wording of questions or preparation of nonverbal materials, unintentional interviewer modification of the question's wording, or the way in which a respondent interprets the question. In the response phase, errors may occur because the respondent gives incorrect information, the interviewer interprets it incorrectly, or recording errors occur. One aspect of this regards form; form-related errors concern psychological orientation toward responding differently to different item formats and include:

1. Leniency: the tendency to rate something too high or too low
2. Central tendency: reluctance to give extreme scores
3. Proximity: giving similar responses to items that occur close to one another (Yu, Albaum, & Swenson, 2003, p. 217)

In the analysis stage, errors of incorrect editing and coding, descriptive summarization, and inference can contribute substantially to measurement error. Measurement error is particularly troublesome for the researcher, since it can arise from many different sources and take on many different forms.

Experimental Error

When an experiment is conducted, the researcher attempts to measure the impact of one or more manipulated independent variables on some dependent variable of interest, while controlling for the influence of all other (i.e., extraneous) variables. Unfortunately, control over all possible extraneous variables is rarely possible. Consequently, what may be measured is not the effect of the independent variables but the effect of the experimental situation itself.

METHODS FOR DEALING WITH POTENTIAL ERRORS

For any research design, recognizing that potential errors exist is one thing, but doing something about them is another matter. There are two basic approaches for handling potential errors:

1. Minimize errors through precision in the research design
2. Measure or estimate the error or its impact

Minimize Error

Two different approaches can be taken to minimize total error. The first uses the research design to minimize errors that may result from each of the individual error components. Much of the material in Chapters 3 through 9 of this book discusses effective research methods, and as such, involves techniques designed to minimize individual errors. This is consistent with our view that research design innately involves error management. However, this approach is often limited by the budget allotted to a project.

The second approach recognizes that individual error components are not necessarily independent of each other. Thus, attempts to minimize one component may lead to an increase in another. Reducing sampling error by increasing sample size, for example, leads to potentially greater non-sampling error. This means that the research designer must trade off errors when developing a research design that minimizes total error. For a fixed project budget, therefore, it may be prudent for the research designer to choose a smaller sample size (which will increase sampling error) if the cost savings by doing this can develop techniques that will reduce nonresponse and/or improve the measurement process. If the reduction in these nonsampling errors exceeds the increase in sampling error, there will be a reduction in total error.

Estimate or Measure Error

Estimating or measuring individual components and total error is not easy, primarily due to the nature of non-sampling errors. There is a body of accepted sampling theory that allows the researcher to estimate sampling error for a probability sample, but nothing comparable exists for non-sampling errors. Consequently, subjective or judgmental estimates must be made.

For individual error components, many diverse procedures can be used to estimate and measure their impact as illustrated in Table 2.1. These are discussed where appropriate in subsequent chapters.

Table 2.1 Selected Methods for Handling Non-Sampling Errors

Type of Error	*Design to Avoid*	*Measure*	*Estimate*
Surrogate Information	Strive for realism	No method of direct measurement, as event has not yet occurred	Use track record of studies, Use surrogate variables
Measurement **1. Instrument induced**	Pretest, alternative wording, alternative positions, etc.	Experiment by using alternative positioning, etc. in a subsample	Estimate will likely be for no bias but some variable error
2. Interviewing-associated (e.g., bias, recording, cheating in telephone and personal interviews)	Select and train interviewers correctly Use same editor of all interviews by one interviewer Use cheater questions Use computer program to analyze for patterns of responses by interviewer	Re-interview subsample using expert interviewer Analysis of variance Use cheater questions Use computer program to analyze for patterns Use interpenetrating sample	Estimate will be for both bias and variable error
3. Response	Randomize response technique Ask for verification checks Cross-check questions Use mail-back technique	Compare with known data	Have interviewer evaluate respondent Estimate will be for both bias and variable error
4. Editing	Prepare editing manual Train editors Require daily return of data	Use master editor to edit subsample	Estimate will be for limited bias, some variable error
5. Coding (text and manually entered data)	Pre-code variables Use coding manual User computer program to clean data	Use master coder to validate subsample	Some bias and variable error
6. Tabulation	Use verification for data entry	Recheck sample of forms	Variable error
7. Analysis	No remedy except competence	Use more competent analyst	
Frame	Use multiple frames	Take subsample of excluded segments	Use compensating weights Use past data
Selection	Make sample element and sample unit the same Use probability sample	Compare with known population	Use compensating weights
Nonresponse	Use callbacks Call at appropriate time Use trained interviewers	Take subsample of nonrespondents	Use Politz-Simmons method Use wave analysis

As a final note, even though the researcher has designed a project to minimize error, it is almost never completely eliminated. Consequently, the error that exists for every project must be estimated or measured. This is recognized for sampling error when probability samples are used, though non-sampling errors typically are ignored. Although estimating or measuring errors is better than ignoring them, there may be times when ignoring non-sampling error may not be that bad. For example, if non-sampling error is viewed as a multiple of sampling error, ignoring non-sampling errors up to an amount equal to one-half of sampling error reduces a .95 confidence level only to .92 (Tull & Albaum, 1973). However, ignoring a non-sampling error equal in amount to sampling error reduces the .95 level to .83.

CHOOSING A RESEARCH DESIGN

The overview of research designs and sources of error just presented should make it apparent that, given a specified problem, many competing designs can provide relevant information. Each design will have an associated expected value of information and incurred cost.

Suppose, for example, that a researcher is assigned to determine the market share of the ten leading brands of energy drinks. There are many possible ways of measuring market share of energy drink brands, including questioning a sample of respondents, observing purchases at a sample of retail outlets, obtaining sales figures from a sample of wholesalers, obtaining sales figures from a sample of retailers and vending machine operators, obtaining tax data, subscribing to a national consumer panel, subscribing to a national panel of retail stores, and, possibly, obtaining data directly from trade association reports or a recent study by some other investigative agency. Though lengthy, this listing is not exhaustive.

The selection of the best design from the alternatives is no different in principle from choosing among the alternatives in making any decision. The associated expected value and cost of information must be determined for each competing design option. If the design is such that the project will yield information for solving more than one problem, the expected value should be determined for all applicable problems and summed. The design with the highest, positive, net expected payoff of research should be selected.

SUMMARY

In this chapter we dealt with a subject of single most importance to the research project: the research design. We described what a research design is, discussed the classes of designs, and examined major sources of marketing information that various designs employ. Finally, we considered the errors that affect research designs.

Presenting these topics as an introduction and overview, we deal with the topics in more depth in the next several chapters. These chapters deal with major sources of marketing information—respondents and experimentation—and the means of obtaining and analyzing research information.

REFERENCES

Anderson, B. F. (1971). *The psychology experiment: An introduction to the scientific method* (2nd ed.). Belmont, CA: Brooks/Cole.

Assael, H., & Keon, J. (1982, Spring). Non-sampling vs. sampling errors in survey research. *Journal of Marketing, 46,* 114–123.

Feldman, J. (1975). Considerations in the use of causal-correlational technique in applied psychology. *Journal of Applied Psychology, 60,* 663–670.

Fern, E. F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research, 19* (February), 1–13.

Fink, A. (2003). *How to design surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing non-response in a national telephone survey. *Public Opinion Quarterly, 64,* 125–148.

Lilien, G., Brown, R., & Searls, K. (1991, January 7). How errors add up. *Marketing News, 25,* 20–21.

Monroe, K. B., & Petroshius, S. M. (n.d.). Developing causal priorities. Unpublished working paper, College of Business, Virginia Polytechnic Institute and State University.

Tomaskovic-Devey, D., Leiter, J., & Thompson, S. (1994). Organizational survey response. *Administrative Science Quarterly, 39,* 439–457.

Tull, D. S., & Albaum, G. S. (1973). *Survey research: A decisional approach.* New York: Intext Educational Publishers.

Yu, J., Albaum, G., & Swenson, M. (2003). Is a central tendency error inherent in the use of semantic differential scales in different cultures? *International Journal of Market Research, 45* (2), 213–228.