SURVEY RESEARCH:

Introduction

The purpose of survey research is to collect primary data—data gathered and assembled specifically for the project at hand. Often research entails asking people—called **respondents**—to provide answers to written or spoken questions. These interviews or questionnaires collect data through the mail, on the telephone, online, or face-to-face. Thus, a survey is defined as a method of collecting primary data based on communication with a representative sample of individuals. Surveys provide a snapshot at a given point in time. The more formal term, **sample survey**, emphasizes that the purpose of contacting respondents is to obtain a representative sample, or subset, of the target population.

Using Surveys

The type of information gathered in a survey varies considerably depending on its objectives. Typically, surveys attempt to describe what is happening or to learn the reasons for a particular business activity.

Identifying characteristics of target markets, measuring customer attitudes, and describing consumer purchase patterns are all common business research objectives. Most business surveys have multiple objectives; few gather only a single type of factual information. In the opening vignette, In-Stat asked questions about product use and desirable features which can help with product development and advertising messages. Geographic, demographic, and media exposure information were also collected to help plan a market segmentation strategy. A study commissioned by eBay provides another example of the information that can be gleaned from survey research. eBay learned that almost 60 percent of respondents receive unwanted gifts, and 15 percent of them had sold an unwanted gift online, suggesting a possible source of demand for eBay's auction services.2 In addition, the survey indicated that selling unwanted gifts online was twice as common among 25- to 34-year-olds. Although consumer surveys are a common form of business research, not all survey research is conducted with the ultimate consumer. Frequently, studies focus on wholesalers, retailers, industrial buyers, or within the organization itself. For example, a survey could be used to determine an organization's commitment to the environment. Also, measuring employee job satisfaction and describing the risk aversion of financial investors may be important survey objectives.

Because most survey research is descriptive research, the term *survey* is most often associated with quantitative findings. Although most surveys are conducted to quantify certain factual information, some aspects of surveys may also be qualitative. In new-product development, a survey often has a qualitative objective of refining product concepts. Stylistic, aesthetic, or functional

Advantages of Surveys

Surveys provide a quick, inexpensive, efficient, and accurate means of assessing information about a population. The examples given earlier illustrate that surveys are quite flexible and, when properly conducted, extremely valuable to the manager. As we discussed in Chapter 1, business research has proliferated in recent years. The growth of survey research is related to the simple idea that to find out what someone thinks, you need to ask them. Over the last 50 years and particularly during the last two decades, survey research techniques and standards have become quite scientific and accurate. When properly conducted, surveys offer managers many advantages.

However, they can also be used poorly when researchers do not follow research principles, such as careful survey and sample design. Sometimes even a well designed and carefully executed survey is not helpful because the results are delivered too late to inform decisions. The disadvantages of specific forms of survey data collection—personal interview, telephone, mail, Internet, and other self-administered formats—are discussed in Chapter 10. However, errors are common to all forms of surveys, so it is appropriate to describe them generally.

Errors in Survey Research

A manager who is evaluating the quality of a survey must estimate its accuracy. Exhibit below outlines the various forms of survey error. They have two major sources: **random sampling error** and **systematic error**.

1-Random Sampling Error

Most surveys try to portray a representative cross-section of a particular target population. Even with technically proper random probability samples, however, statistical errors will occur because of chance variation in the elements selected for the sample. These statistical problems are unavoidable without very large samples (>400). However, the extent of **random sampling error** can be estimated.

2-Systematic Error

The other major source of survey error, **systematic error**, results from some imperfect aspect of the research design or from a mistake in the execution of the research. Because systematic errors include all sources of error other than those introduced by the random sampling procedure, these errors or biases are also called *nonsampling errors.* A **sample bias** exists when the results of a sample show a persistent tendency to deviate in one direction from the true value of the population parameter. The many sources of error that in some way systematically influence answers can be divided into two general categories: **respondent error** and **administrative error**.

i-Respondent Error

Surveys ask people for answers. If people cooperate and give truthful answers, a survey will likely accomplish its goal. If these conditions are not met, **nonresponse error or response bias**, the two major categories of respondent error, may cause sample bias.

a-Nonresponse Error

Few surveys have 100 percent response rates. In fact, surveys with relatively low response rates may still accurately reflect the population of interest. However, a researcher who obtains a 1 percent response to a five-page e-mail questionnaire concerning various brands of spark plugs may face a serious problem. To use the results, the researcher must believe that consumers who responded to the questionnaire are representative of all consumers, including those who did not respond. The statistical differences between a survey that includes only those who responded and a survey that also included those who failed to respond are referred to as **nonresponse error**. This problem is especially acute in mail and Internet surveys, but nonresponse also threatens telephone and face-to-face interviews.

People who are not contacted or who refuse to cooperate are called **nonrespondents**. A nonresponse occurs if no one answers the phone at the time of both the initial call and any subsequent callbacks. The number of **no contacts** in telephone survey research has been increasing because of the proliferation of answering machines and growing use of caller ID to screen telephone calls. The respondent who is not at home when called or visited should be scheduled to be interviewed at a different time of day or on a different day of the week. **Refusals** occur when people are unwilling to participate in the research. A parent who must juggle the telephone and a half-diapered child and refuses to participate in the survey because he or she is too busy also is a nonresponse. After receiving a refusal from a potential respondent, an interviewer can do nothing other than be polite.

With a mail survey, the researcher never really knows whether a nonrespondent actually received the survey, has refused to participate, or is just indifferent. Researchers know that those who are most involved in an issue are more likely to respond to a mail survey. **Self-selection bias** is a problem that frequently plagues self-administered questionnaires. In a restaurant, for example, a customer on whom a waiter spilled soup, a person who was treated to a surprise dinner, or others

who feel strongly about the service are more likely to complete a self-administered questionnaire left at the table than individuals who are indifferent about the restaurant. Self-selection biases distort surveys because they overrepresent extreme positions while underrepresenting responses from those who are indifferent. Several techniques will be discussed later for encouraging respondents to reply to mail and Internet surveys.

Comparing the demographics of the sample with the demographics of the target population is one means of inspecting for possible biases in response patterns. If a particular group, such as older citizens, is underrepresented or if any potential biases appear in a response pattern, additional efforts should be made to obtain data from the underrepresented segments of the population. For example, telephone surveys may be used instead of mail surveys or personal interviews may be used instead of telephone interviews in an attempt to increase participation of underrepresented segments.

b-RESPONSE BIAS

A **response bias** occurs when respondents tend to answer questions with a certain slant. People may consciously or unconsciously misrepresent the truth. If a distortion of measurement occurs because respondents' answers are falsified or misrepresented, either intentionally or inadvertently, the resulting sample bias will be a response bias. When researchers identify response bias, they should include a corrective measure.

Deliberate Falsification

Occasionally people deliberately give false answers. It is difficult to assess why people knowingly misrepresent answers. A response bias may occur when people misrepresent answers to appear intelligent, conceal personal information, avoid embarrassment, and so on. For example, respondents may be able to remember the total amount of money spent grocery shopping, but they may forget the exact prices of individual items that they purchased. Rather than appear ignorant or unconcerned about prices, they may provide their best estimate and not tell the truth—namely, that they cannot remember. Sometimes respondents become bored with the interview and provide answers just to get rid of the interviewer. At other times respondents try to appear well informed by providing the answers they think are expected of them. On still other occasions, they give answers simply to please the interviewer. One explanation for conscious and deliberate misrepresentation of facts is the so-called average-person hypothesis. Individuals may prefer to be viewed as average, so they alter their responses to conform more closely to their *perception* of the average person. Average-person effects have been found in response to questions about such topics as savings account balances, car prices, voting behavior, and hospital stays.

Unconscious Misrepresentation

Even when a respondent is consciously trying to be truthful and cooperative, response bias can arise from the question format, the question content, or some other stimulus. For example, bias can be introduced by the situation in which the survey is administered. The results of two in- flight surveys concerning aircraft preference illustrate this point. Passengers flying on B-747s preferred B-747s to L-1011s (74 percent versus 19 percent), while passengers flying on L-1011s preferred L-1011s to B-747s (56 percent versus 38 percent). Managers may be tempted to conclude that the results demonstrate a preference for B-747s over L-1011s. But perhaps respondents were influenced by other factors besides the airplane or had not experience flying in the other type of plane. Respondents' satisfaction scores may simply be a simple response to their overall satisfaction with the flying experience. Also, airlines have fleets that consist predominantly of one brand of aircraft. Thus, the data appearing to support Boeing may really be showing greater satisfaction for the airlines that happen to be flying Boeing and would have higher satisfaction no matter what planes were in their fleet.

Types of Response Bias

Response bias falls into four specific categories: acquiescence bias, extremity bias, interviewer bias, and social desirability bias. These categories overlap and are not mutually exclusive. A single biased answer may be distorted for many complex reasons, some distortions being deliberate and some being unconscious misrepresentations.

- **Acquiescence Bias:-** Some respondents are very agreeable. They seem to agree to practically every statement they are asked about. A tendency to agree (or disagree) with all or most questions is known as **acquiescence bias**. This bias is particularly prominent in new- product research. Questions about a new-product idea generally elicit some acquiescence bias because respondents give positive connotations to most new ideas. For example, consumers responded favorably to survey questions about pump baseball gloves (the pump inserts air into the pocket of the glove, providing more cushioning). However, when these expensive gloves hit the market, they sat on the shelves. When conducting new-product research, researchers should recognize the high likelihood of acquiescence bias.
- **Extremity Bias:-** Some individuals tend to use extremes when responding to questions. For example, they may choose only "1" or "10" on a ten-point scale. Others consistently refuse to use extreme positions and tend to respond more neutrally—"I never give a 10 because nothing is really perfect." Response styles vary from person to person, and extreme responses may cause an **extremity bias** in the data.
- **Interviewer Bias:-** Response bias may arise from the interplay between interviewer and respondent. If the interviewer's presence influences respondents to give untrue or modified answers, the survey will be marred by **interviewer bias**. Many homemakers and retired people welcome an interviewer's visit as a break in routine activities. Other respondents may give answers they believe will please the interviewer rather than the truthful responses. Respondents may wish to appear intelligent and wealthy—of course they read *Scientific American* rather than *Playboy*! The interviewer's age, sex, style of dress, tone of voice, facial expressions, or other nonverbal characteristics may have some influence on a respondent's answers. If an interviewer smiles and makes a positive statement after a respondent's answers, the respondent will be more likely to give similar responses. In a research study on sexual harassment against saleswomen, male interviewers might not yield as candid responses from saleswomen as female interviewers would. Many interviewers, contrary to instructions, shorten or rephrase needs. This potential influence on responses can be avoided to some extent if interviewers receive training and supervision that emphasize the necessity of appearing neutral. If interviews go on too long, respondents may feel that time is being wasted. They may answer as abruptly as possible with little forethought.
- **Social Desirability Bias:- Social desirability bias** may occur either consciously or unconsciously because the respondent wishes to create a favorable impression or save face in the presence of an interviewer. Incomes may be inflated, education overstated, or perceived respectable answers given to gain prestige. In contrast, answers to questions that seek factual information or responses about matters of public knowledge (zip code, number of children, and so on) usually are quite accurate. An interviewer's presence may increase a respondent's tendency to give inaccurate answers to sensitive questions such as "Did you vote in the last election?" or "Do you have termites or roaches in your home?" or "Do you color your hair?" Social desirability bias is especially significant in the case of research that addresses sensitive or personal topics, including respondents' sexual behavior. A group of researchers recently evaluated responses to questions about homosexual sexual activity, collected by the National Opinion Research Center's long running General Social Survey. The researchers found that over time, as attitudes toward homosexual conduct have softened, the frequency of repeated female sexual contacts increased.

ii- Administrative Error

The result of improper administration or execution of the research task is called an **administrative error**. Administrative errors are caused by carelessness, confusion, neglect, omission, or some other blunder. Four types of administrative error are

- data-processing error,
- sample selection error,
- interviewer error,
- interviewer cheating.

DATA PROCESSING ERROR

Processing data by computer, like any arithmetic or procedural process, is subject to error because data must be edited, coded, and entered into the computer by people. The accuracy of data processed by computer depends on correct data entry and programming. **Data-processing error** can be minimized by establishing careful procedures for verifying each step in the dataprocessing stage.

SAMPLE SELECTION ERROR

Many kinds of error involve failure to select a representative sample. **Sample selection error** is systematic error that results in an unrepresentative sample because of an error in either the sample design or the execution of the sampling procedure. Executing a sampling plan free of procedural error is difficult. A firm that selects its sample from the phone book will have some systematic error, because those with only cell phones or with unlisted numbers are not included. Stopping respondents during daytime hours in shopping centers largely excludes working people or those who primarily shop by mail, Internet, or telephone. In other cases, researchers interview the wrong person. Consider a political pollster who uses random-digit dialing to select a sample, rather than a list of registered voters. Unregistered 17-year-olds may be willing to give their opinions, but they are the wrong people to ask because they cannot vote.

INTERVIEWER ERROR

Interviewers' abilities vary considerably. **Interviewer error** is introduced when interviewers record answers but check the wrong response or are unable to write fast enough to record answers verbatim. Also, selective perception may cause interviewers to misrecord data that do not support their own attitudes and opinions.

INTERVIEWER CHEATING

Interviewer cheating occurs when an interviewer falsifies entire questionnaires or fills in answers to questions that have been intentionally skipped. Some interviewers cheat to finish an interview as quickly as possible or to avoid questions about sensitive topics. Often interviewers are paid by the completed survey, so you can see the motivation to complete a survey that is left with some questions unanswered. If interviewers are suspected of faking questionnaires, they should be told that a small percentage of respondents will be called back to confirm whether the initial interview was actually conducted. This practice should discourage interviewers from cheating. The term *curb-stoning* is sometimes used to refer to interviewers filling in responses for respondents that do not really exist.

Rule-of-Thumb Estimates for Systematic Error

The techniques for estimating systematic, or nonsampling, error are less precise than many sample statistics. Researchers have established conservative rules of thumb based on experience to estimate systematic error. In the case of consumer research, experienced researchers might determine that only a certain percentage of people who say they will definitely buy a new product actually do so. Evidence for a mere-measurement effect (see the Research Snapshot on the next page) suggests that

in some situations, researchers might conclude that respondents' own buying behavior will exaggerate overall sales. Thus, researchers often present actual survey findings *and* their interpretations of estimated purchase response based on estimates of nonsampling error. For example, one pay-per-view cable TV company surveys geographic areas it plans to enter and estimates the number of people who indicate they will subscribe to its service. The company knocks down the percentage by a "ballpark 10 percent" because experience in other geographic areas has indicated that there is a systematic upward bias of 10 percent on this intentions question.

Classifying Survey Research Methods

Now that we have introduced some advantages and disadvantages of surveys in general, we turn to a discussion of classification of surveys according to several criteria. Surveys may be classified based on the method of communication, the degrees of structure and disguise in the questionnaire, and the time frame in which the data are gathered (temporal classification).

Structured/Unstructured and Disguised/ Undisguised Questionnaires

In designing a questionnaire (or an *interview schedule*), the researcher must decide how much structure or standardization is needed. A **structured question** limits the number of allowable responses. For example, the respondent may be instructed to choose one alternative response such as "under 18," "18–35," or "over 35" to indicate his or her age. An **unstructured question** does not restrict the respondent's answers. An open-ended, unstructured question such as "Why do you shop at Wal-Mart?" allows the respondent considerable freedom in answering. The researcher must also decide whether to use **undisguised questions** or **disguised questions**. A straightforward, or undisguised, question such as "Do you have dandruff problems?" assumes that the respondent is willing to reveal the information. However, researchers know that some questions are threatening to a person's ego, prestige, or self-concept. So, they have designed a number of indirect techniques of questioning to disguise the purpose of the study. Questionnaires can be categorized by their degree of structure and degree of disguise. For example, interviews in exploratory research might use *unstructured-disguised* questionnaires. The projective techniques discussed in Chapter 7 fall into this category. Other classifications are *structured-undisguised, unstructured-undisguised,* and *structured-disguised.* These classifications have two limitations: First, the degree of structure and the degree of disguise vary; they are not clear-cut categories. Second, most surveys are hybrids, asking both structured and unstructured questions. Recognizing the degrees of structure and disguise necessary to meet survey objectives will help in the selection of the appropriate communication medium for conducting the survey.

Temporal Classification

Although most surveys are for individual research projects conducted only once over a short time period, other projects require multiple surveys over a long period. Thus, surveys can be classified on a temporal basis.

CROSS SECTIONAL STUDIES

Do you make New Year's resolutions? A Harris Interactive survey conducted in November 2008 indicates that women (74 percent) are more likely than men (58 percent) to actually make a New Year's resolution. However, more men than women "always or often keep their resolutions" (22 percent of men compared to 14 percent of women).12 This was a **cross-sectional study** because it collected the data at a single point in time. That is, the survey asked people to reflect on their past behavior, rather than ask them if they made a resolution, then follow up a year later to see if the resolution was kept. Such a study samples various segments of the population to investigate relationships among variables by cross-tabulation. Most business research surveys fall into this category. We can think of cross-sectional studies as taking a snapshot of the current situation. The typical method of analyzing a cross-sectional survey is to divide the sample into appropriate subgroups. For example, if a winery expects income levels to influence attitudes toward wines, the

data are broken down into subgroups based on income and analyzed to reveal similarities or differences among the income subgroups. If a manager thinks that length of time an employee has been with the organization will influence their attitudes toward corporate policies, employees might be broken into different groups based on tenure (e.g., less than 5 years, 5–9 years, 10–14 years, and 15 years or more) so their attitudes can be examined.

LONGITUDINAL STUDIES

In a **longitudinal study** respondents are questioned at multiple points in time. The purpose of longitudinal studies is to examine continuity of response and to observe changes that occur over time. Many syndicated polling services, such as Gallup, conduct regular polls. For example, the Bureau of Labor Statistics conducts the National Longitudinal Survey of Youth, interviewing the same sample of individuals repeatedly since 1979. (Respondents, who were "youth" at the beginning of the study, are now in their 40s.) Research scientist Jay Zagorsky recently analyzed the longitudinal data from that study to determine that those who married and stayed with their spouse accumulated almost twice as much wealth as single and divorced people in the study.13 The Yankelovich MONITOR has been tracking American values and attitudes for more than 30 years. This survey is an example of a longitudinal study that uses successive samples; its researchers survey several different samples at different times. Longitudinal studies of this type are sometimes called *cohort studies,* because similar groups of people who share a certain experience during the same time interval (cohorts) are expected to be included in each sample. Exhibit below illustrates the results of a longitudinal study by Harris Interactive, which since 1966 has been asking five questions related to powerlessness and isolation to create an "alienation index."

Consumer Panel

A longitudinal study that gathers data from the same sample of individuals or households over time is called a **consumer panel**. Consider the packaged-goods marketer that wishes to learn about brand-switching behavior. A consumer panel that consists of a group of people who record their purchasing habits in a diary over time will provide the manager with a continuous stream of information about the brand and product class. Diary data that are recorded regularly over an extended period enable the researcher to track repeat-purchase behavior and changes in purchasing habits that occur in response to changes in price, special promotions, or other aspects of business strategy. Panel members may be contacted by telephone, in a personal interview, by mail questionnaire, or by e-mail. Typically respondents complete media exposure or purchase diaries and mail them back to the survey organization. If the panel members have agreed to field test new products, face-to-face or telephone interviews may be required. The nature of the problem dictates which communication method to use. Because establishing and maintaining a panel is expensive, panels often are managed by contractors who offer their services to many organizations. A number of commercial firms, such as National Family Opinion (NFO), Inc., Market Research Corporation of America, and Consumer Mail Panels, Inc., specialize in maintaining consumer panels. In recent years Internet panels have grown in popularity. Because clients of these firms need to share the expenses with other clients to acquire longitudinal data at a reasonable cost, panel members may be asked questions about a number of different issues.

The first questionnaire a panel member is asked to complete typically includes questions about product ownership, product usage, pets, family members, and demographic data. The purpose of such a questionnaire is to gather the behavioral and demographic data that will be used to identify heavy buyers, difficult-to-reach customers, and so on for future surveys. Individuals who serve as members of consumer panels usually are compensated with cash, attractive gifts, or the chance to win a sweepstakes. Marketers whose products are purchased by few households find panels an economical means of reaching respondents who own their products. A two-stage process typically is used. A panel composed of around 15,000 households can be screened with a one-question statement attached to another project. For example, a question in an NFO questionnaire screens for ownership of certain uncommon products, such as snowmobiles and motorcycles. This information is stored in a database. Then households with the unusual item can be sampled again with a longer questionnaire.

Total Quality Management and Customer Satisfaction Surveys

Total quality management is a business strategy that emphasizes market-driven quality as a top priority. Total quality management involves implementing and adjusting the firm's business activities to assure customers' satisfaction with the quality of goods and services. Many U.S. organizations adopted total quality management in the 1980s when an increase in high-quality foreign competition challenged their former dominance. Today companies continue to recognize the need for total quality management programs. Executives and production workers are sometimes too far removed from the customer. Companies need a means to bridge this gap with feedback about quality of goods and services. This means conducting research. In an organization driven by the quality concept, business research plays an important role in the management of total product quality.

What Is Quality?

Organizations used to define quality by engineering standards. Most companies no longer see quality that way. Some managers say that having a quality product means that the good or service conforms to consumers' requirements and that the product is acceptable. Effective executives who subscribe to a total quality management philosophy, however, believe that the product's quality must go beyond acceptability for a given price range. Rather than merely being relieved that nothing went wrong, consumers should experience some delightful surprises or reap some unexpected benefits. In other words, quality assurance is more than just meeting minimum standards. The level of quality is the degree to which a good or service truly is seen as good or bad.

Internal and External Customers

Organizations that have adopted the total quality management philosophy believe that a focus on customers must include more than external customers. Like Arbor, Inc., they believe that everyone in the organization has customers:

Every person, in every department, and at every level, has a customer. The customer is anyone to whom an individual provides service, information, support, or product. The customer may be another employee or department (internal) or outside the company (external).

Total quality management programs work most effectively when every employee knows exactly who his or her customers are and what output internal and external customers expect. Also, it is important to know how customers perceive their needs are being met. All too often differences between perceptions and reality are not understood.

Implementing Total Quality Management

Implementing a total quality management program requires considerable survey research. A firm must routinely ask customers to rate it against its competitors. It must periodically measure employee knowledge, attitudes, and expectations. It must monitor company performance against benchmark standards. It must determine whether customers found any delightful surprises or major disappointments. In other words, a total quality management strategy expresses the conviction that to improve quality, an organization must regularly conduct surveys to evaluate quality improvement.

Table below illustrates the total quality management process. The exhibit shows that overall tracking of quality improvement requires longitudinal research. The process begins with a *commitment and exploration stage,* during which management makes a commitment to total quality assurance and researchers explore external and internal customers' needs and beliefs. The research must discover what product features customers value, what problems customers are having with the product, what aspects of product operation or customer service have disappointed customers, what the company is doing right, and what the company may be doing wrong.

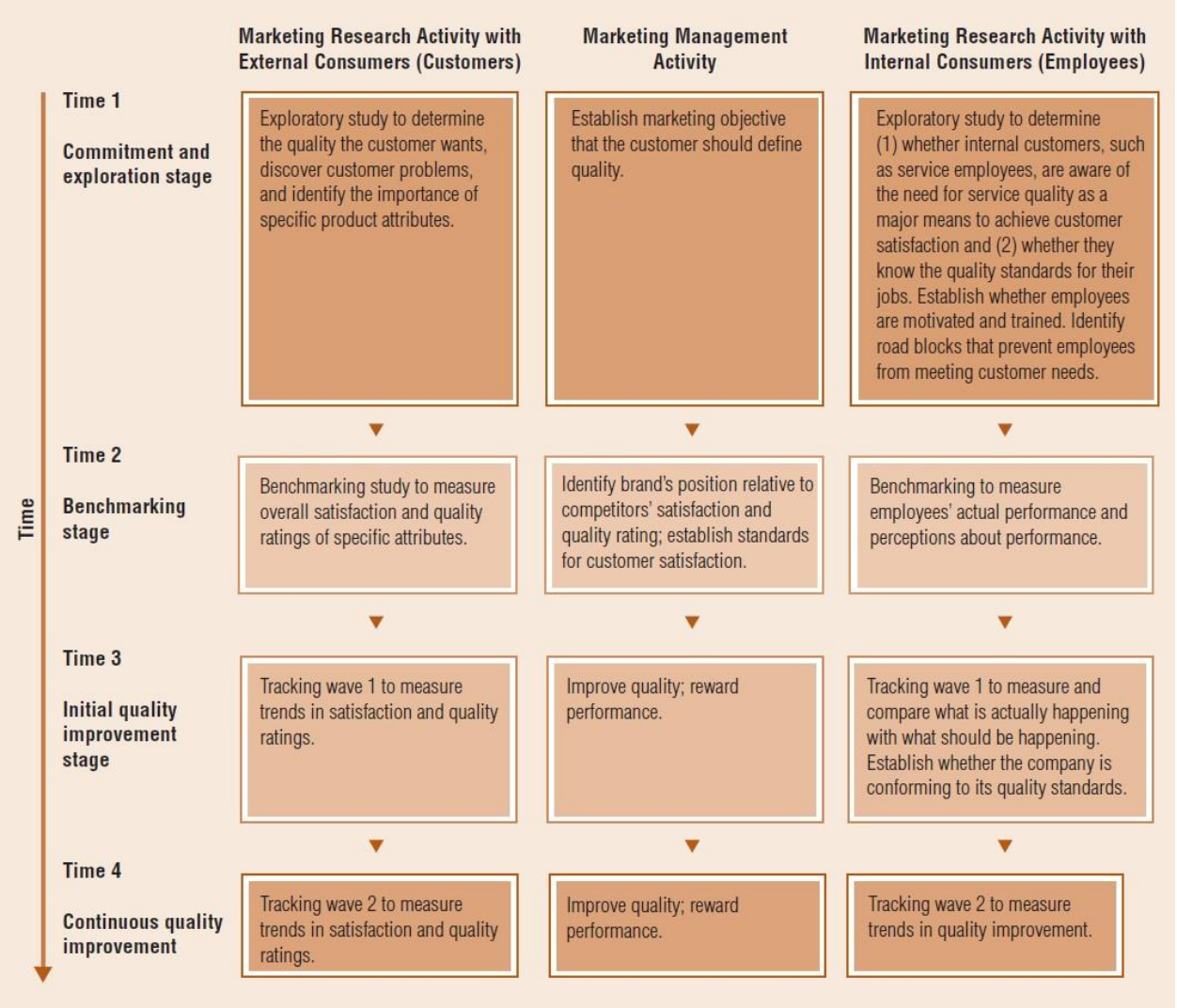


Table- Longitudinal Research for Total Quality Management

After internal and external customers' problems and desires have been identified, the *benchmarking stage* begins. Research must establish quantitative measures that can serve as benchmarks or points of comparison against which to evaluate future efforts. The surveys must establish initial measures of overall satisfaction, of the frequency of customer problems, and of quality ratings for specific attributes. Researchers must identify the company's or brand's position relative to competitors' quality positions. For example, when Anthony Balzarini became food-service manager at Empire Health Services in Spokane, Washington, he became responsible for serving meals to the patients of the company's two hospitals, plus retail food service (sales to visitors and employees who eat in the hospitals). He began tracking quality according to several measures, including satisfaction scores on patient surveys and sales volume and revenue on the retail side. Sales measurements include comparing the average sale with other locations, including restaurants, in the Spokane area.

The *initial quality improvement* stage establishes a quality improvement process within the organization. Management and employees must translate quality issues into the internal vocabulary of the organization. The company must establish performance standards and expectations for improvement. For Balzarini, this stage included training food-service employees in providing patient service. He began holding meetings twice daily to identify any problems to be resolved. Managers were each assigned to one floor of the hospital and charged with building a close working relationship with the nursing staff there. They are expected to visit their floor every week and conduct 15 interviews with patients to learn about what they like and dislike. On the retail side, the manager is expected to revise menus every 12 weeks to offer more variety. Waste is literally weighed and categorized to identify

which types of food are rejected by patients and customers.

After managers and employees have set quality objectives and implemented procedures and standards, the firm continues to track satisfaction and quality ratings in successive waves. The purpose of tracking wave 1 is to measure trends in satisfaction and quality ratings. Business researchers determine whether the organization is meeting customer needs as specified by quantitative standards. At one of Empire's two hospitals, one of the food-service managers learned that a patient on a liquid diet disliked the broth he was being served. An investigation showed that the recipe had been changed, and a taste test confirmed that the original recipe was superior, so the hospital switched back to the original recipe.

The next stage, *continuous quality improvement*, consists of many consecutive waves with the same purpose—to improve over the previous period. Continuous quality improvement requires that management allow employees to initiate problem solving without a lot of red tape. Employees should be able to initiate proactive communications with consumers. In tracking wave 2, management compares results with those of earlier stages. Quality improvement management continues. At Empire, improvements have been reflected in rising patient satisfaction scores and growing sales in retail operations.

Management must also reward performance. At Empire, Balzarini set up a program called "You Rock." Any employee who observes an excellent action by another employee, beyond mere job requirements, acknowledges the good work with a card awarding points redeemable in the hospitals' retail areas. Balzarini also sends weekly thank-you cards to workers who showed outstanding performance.

Table above shows that total quality management programs measure performance against *customers'* standards—not against standards determined by quality engineers within the company. All changes within the organization are oriented toward improvement of customers' perceptions of quality. The exhibit indicates the need for integration of establishing consumer requirements, quantifying benchmark measures, setting objectives, conducting research studies, and making adjustments in the organization to improve quality. Continuous quality improvement is an ongoing process. The activities outlined in Exhibit above work for providers of both goods and services. However, service products and customer services offered along with goods have some distinctive aspects. We will first discuss the quality of goods and then consider the quality of services.

SURVEY RESEARCH:COMMUNICATING WITH RESPONDENTS

Introduction

During most of the twentieth century, obtaining survey data involved inviting individuals to answer questions asked by human interviewers (interviews) or questions they read themselves (questionnaires). Interviewers communicated with respondents face-to-face or over the telephone, or respondents filled out self-administered paper questionnaires, which were typically distributed by mail. These media for conducting surveys remain popular with business researchers. However, as the preceding vignette suggests, digital technology is having a profound impact on society in general and on business research in particular. Its greatest impact is in the creation of new forms of communications media.

Interviews as Interactive Communication

When two people engage in a conversation, human interaction takes place. Human interactive media are a personal form of communication. One human being directs a message to and interacts with another individual (or a small group). When most people think of interviewing, they envision two people engaged in a face-to-face dialogue or a conversation on the telephone.

Electronic interactive media allow researchers to reach a large audience, personalize individual messages, and interact using digital technology. To a large extent, electronic interactive media are controlled by the users themselves. No other human need be present. Survey respondents today are not passive audience members. They are actively involved in a two-way communication using electronic interactive media.

The Internet is radically altering many organizations' research strategies, providing a prominent example of the new electronic interactive media. Consumers determine what information they will be exposed to by choosing what sites to visit and by blocking or closing annoying pop-up ads. Electronic interactive media also include CDROM and DVD materials, touch-tone telephone systems, touch-screen interactive kiosks in stores, and other forms of digital technology.

Noninteractive Media

The traditional questionnaire received by mail and completed by the respondent does not allow a dialogue or an exchange of information providing immediate feedback. So, from our perspective, self-administered questionnaires printed on paper are noninteractive. This fact does not mean that they are without merit, just that this type of survey is less flexible than surveys using interactive communication media. Each technique for conducting surveys has merits and shortcomings. The purpose of this chapter is to explain when researchers should use different types of surveys. The chapter begins with a discussion of surveys that use live interviews. Then we turn to noninteractive, selfadministered questionnaires. Finally, we explain how the Internet and digital technology are dramatically changing survey research.

Personal Interviews

To conduct interviews, the researcher may communicate with individuals in person by going door-to-door or intercepting them in shopping malls, or interviews may take place over the telephone. Traditionally, researchers have recorded interview results using paper and pencil, but computers are increasingly supporting survey research. In this section, we examine the general characteristics of face-to-face personal interviews, then compare the characteristics of door-to- door personal interviews and personal interviews conducted in shopping malls. The next section examines telephone interviews.

Gathering information through face-to-face contact with individuals goes back many years. Periodic censuses were used to set tax rates and aid military conscription in the ancient empires of Egypt and Rome. During the Middle Ages, the merchant families of Fugger and Rothschild prospered in part because their far-flung organizations enabled them to get information before their competitors could. Today, survey researchers typically present themselves in shopping centers and street corners throughout the United States and announce, "Good afternoon, my name is_____. I am with _____ Company, and we are conducting a survey on_____." A **personal interview** is a form of direct communication in which an interviewer asks respondents questions face-to-face. This versatile and flexible method is a two-way conversation between interviewer and respondent.

Advantages of Personal Interviews

Business researchers find that personal interviews offer many unique advantages. One of the most important is the opportunity for detailed feedback.

- **OPPORTUNITY FOR FEEDBACK:** Personal interviews, similar to those mentioned in the Research Snapshot on the next page, provide the opportunity for feedback and clarification. For example, if a consumer is reluctant to provide sensitive information, the interviewer may offer reassurance that his or her answers will be strictly confidential. Personal interviews offer the lowest chance that respondents will misinterpret questions, because an interviewer who senses confusion can clarify the instruction or questions. Circumstances may dictate that at the conclusion of the interview, the respondent be given additional information concerning the

purpose of the study. This clarification is easily accomplished with a personal interview. If the feedback indicates that some question or set of questions is particularly confusing, the researcher can make changes that make the questionnaire easier to understand.

- **PROBING COMPLEX ANSWERS**: Another important characteristic of personal interviews is the opportunity to follow up by probing. If a respondent's answer is too brief or unclear, the researcher may request a more comprehensive or clearer explanation. In probing, the interviewer asks for clarification with standardized questions such as "Can you tell me more about what you had in mind?". Although interviewers are expected to ask questions exactly as they appear on the questionnaire, probing allows them some flexibility. Depending on the research purpose, personal interviews vary in the degree to which questions are structured and in the amount of probing required. The personal interview is especially useful for obtaining unstructured information. Skilled interviewers can handle complex questions that cannot easily be asked in telephone or mail surveys.
- **LENGTH OF INTERVIEW:** If the research objective requires an extremely lengthy questionnaire, personal interviews may be the only option. A general rule of thumb on mail surveys is that they should not exceed six pages, and telephone interviews typically last less than ten minutes. In contrast, a personal interview can be much longer, perhaps an hour and a half, as was the case for the U.S. National Adult Literacy Assessment. However, the longer the interview, no matter what the form, the more the respondent should be compensated for their time and participation. Researchers should also be clear about how long participation should take in the opening dialog requesting participation. Online surveys should include a completion meter that shows the progress a respondent has made toward completing the task.
- **COMPLETENESS OF QUESTIONNAIRE:** The social interaction between a well-trained interviewer and a respondent in a personal interview increases the likelihood that the respondent will answer all the items on the questionnaire. The respondent who grows bored with a telephone interview may terminate the interview at his or her discretion simply by hanging up the phone. Self-administration of a mail questionnaire requires even more effort by the respondent. Rather than write lengthy responses, the respondent may fail to complete some of the questions. **Item nonresponse**—failure to provide an answer to a question—is least likely to occur when an experienced interviewer asks questions directly.
- **PROPS AND VISUAL AIDS:** Interviewing respondents face-to-face allows the investigator to show them new product samples, sketches of proposed advertising, or other visual aids. When Lego Group wanted to introduce new train model sets for its famous building bricks, the company targeted adults who build complex models with its product. The company invited adults who were swapping ideas at the Lego Web site to visit the New York office, where they viewed ideas and provided their opinions. The respondents wound up rejecting all the company's ideas, but they suggested something different: the Santa Fe Super Chief set, which sold out within two weeks, after being advertised only by enthusiastic word of mouth. This research could not have been done in a telephone interview or mail survey. Research that uses visual aids has become increasingly popular with researchers who investigate film concepts, advertising problems, and moviegoers' awareness of performers. Research for movies often begins by showing respondents videotapes of the prospective cast. After the movie has been produced, film clips are shown and interviews conducted to evaluate the movie's appeal, especially which scenes to emphasize in advertisements.
- **HIGH PARTICIPATION:** Although some people are reluctant to participate in a survey, the presence of an interviewer generally increases the percentage of people willing to complete the interview. Respondents typically are required to do no reading or writing—all they have to do is talk. Many people enjoy sharing information and insights with friendly and sympathetic interviewers. People are often more hesitant to tell a person "no" face-to-face than they are over the phone or through some impersonal contact.

Disadvantages of Personal Interviews

Personal interviews also have some disadvantages. Respondents are not anonymous and as a result may be reluctant to provide confidential information to another person. Suppose a survey asked top executives, "Do you see any major internal instabilities or threats (people, money, material, and so on) to the achievement of your marketing objectives?" Many managers may be reluctant to answer this sensitive question honestly in a personal interview in which their identities are known.

- **INTERVIEWER INFLUENCE:** Some evidence suggests that demographic characteristics of the interviewer influence respondents' answers. For example, one research study revealed that male interviewers produced larger amounts of interviewer variance than female interviewers in a survey in which 85 percent of the respondents were female. Older interviewers who interviewed older respondents produced more variance than other age combinations, whereas younger interviewers who interviewed younger respondents produced the least variance. Differential interviewer techniques may be a source of bias. The rephrasing of a question, the interviewer's tone of voice, and the interviewer's appearance may influence the respondent's answer. Consider the interviewer who has conducted 100 personal interviews. During the next one, he or she may lose concentration and either selectively perceive or anticipate the respondent's answer. The interpretation of the response may differ somewhat from what the respondent intended. Typically, the public thinks of the person who does marketing research as a dedicated scientist. Unfortunately, some interviewers do not fit that ideal. Considerable interviewer variability exists. Cheating is possible; interviewers may cut corners to save time and energy, faking parts of their reports by dummying up part, or all, of the questionnaire. Control over interviewers is important to ensure that difficult, embarrassing, or time-consuming questions are handled properly.

- **LACK OF ANONYMITY OF RESPONDENT:** Because a respondent in a personal interview is not anonymous and may be reluctant to provide confidential information to another person, researchers often spend considerable time and effort to phrase sensitive questions to avoid social desirability bias. For example, the interviewer may show the respondent a card that lists possible answers and ask the respondent to read a category number rather than be required to verbalize sensitive answers.

- **COST:** Personal interviews are expensive, generally substantially more costly than mail, Internet, or telephone surveys. The geographic proximity of respondents, the length and complexity of the questionnaire, and the number of people who are nonrespondents because they could not be contacted (not-at-homes) will all influence the cost of the personal interview.

Door-to-Door Interviews and Shopping Mall Intercepts

Personal interviews may be conducted at the respondents' homes or offices or in many other places. Increasingly, personal interviews are being conducted in shopping malls. Mall intercept interviews allow many interviews to be conducted quickly. Often, respondents are intercepted in public areas of shopping malls and then asked to come to a permanent research facility to taste new food items or to view advertisements. The locale for the interview generally influences the participation rate, and thus the degree to which the sample represents the general population.

- ### Door to Door Interviews

The presence of an interviewer at the door generally increases the likelihood that a person will be willing to complete an interview. Because **door-to-door interviews** increase the participation rate, they provide a more representative sample of the population than mail questionnaires. For example, response rates to mail surveys are substantially lower among Hispanics whether the questionnaire is printed in English or Spanish.4 People who do not have telephones, who have unlisted telephone

numbers, or who are otherwise difficult to contact may be reached using door- to-door interviews. However, door-to-door interviews may underrepresent some groups and overrepresent others based on the geographic areas covered. Door-to-door interviews may exclude individuals who live in multiple-dwelling units with security systems, such as high-rise apartment dwellers, or executives who are too busy to grant personal interviews during business hours. Other people, for security reasons, simply will not open the door when a stranger knocks. As seen in the Research Snapshot on the next page, elderly adults, or people in retirement dwellings, may also be excluded. Telephoning an individual in one of these subgroups to make an appointment may make the total sample more representative. However, obtaining a representative sample of this security-conscious subgroup based on a listing in the telephone directory may be difficult. For these reasons, door-to-door interviews are becoming a thing of the past.

- **Callbacks**

When a person selected to be in the sample cannot be contacted on the first visit, a systematic procedure is normally initiated to call back at another time. **Callbacks**, or attempts to recontact individuals selected for the sample, are the major means of reducing nonresponse error. Calling back a sampling unit is more expensive than interviewing the person the first time around, because subjects who initially were not at home generally are more widely dispersed geographically than the original sample units. Callbacks in door-to-door interviews are important because not-at home individuals (for example, working parents) may systematically vary from those who *are* at home (nonworking parents, retired people, and the like).

- ## Mall Intercept Interviews

Personal interviews conducted in shopping malls are referred to as **mall intercept interviews**, or *shopping center sampling.* Interviewers typically intercept shoppers at a central point within the mall or at an entrance. The main reason mall intercept interviews are conducted is because their costs are lower. No travel is required to the respondent's home; instead, the respondent comes to the interviewer, and many interviews can be conducted quickly in this way.

A major problem with mall intercept interviews is that individuals usually are in a hurry to shop, so the incidence of refusal is high—typically around 50 percent. Yet it is standard practice for many commercial research companies, who conduct more personal interviews in shopping malls than it conducts door-to-door.

In a mall interview, the researcher must recognize that he or she should not be looking for a representative sample of the total population. Each mall has its own target market's characteristics, and there is likely to be a larger bias than with careful household probability sampling. However, personal interviews in shopping malls are appropriate when the target group is a special market segment such as the parents of children of bike-riding age. If the respondent indicates that he or she has a child of this age, the parent can then be brought into a rented space and shown several bikes. The mall intercept interview allows the researcher to show large, heavy, or immobile visual materials, such as a television commercial. A mall interviewer can give an individual a product to take home to use and obtain a commitment that the respondent will cooperate when recontacted later by telephone. Mall intercept interviews are also valuable when activities such as cooking and tasting of food must be closely coordinated and timed to follow each other. They may also be appropriate when a consumer durable product must be demonstrated. For example, when videocassette recorders and DVD players were innovations in the prototype stage, the effort and space required to set up and properly display these units ruled out in-home testing.

Telephone Interviews

For several decades, landline **telephone interviews** have been the mainstay of commercial survey research. The quality of data obtained by telephone is potentially comparable to the quality of data collected face-to-face. Respondents are more willing to provide detailed and reliable information on a variety of personal topics over the phone while in the privacy of their own homes than when answering questions face-to-face.

In-home phone surveys are still considered capable of providing fairly representative samples of the U.S. population. However, the "no-call" legislation dating back to the middle of the first decade of the twenty-first century has limited this capability somewhat. Business researchers cannot solicit information via phone numbers listed on the do-not-call registry. Thus, to the extent that consumers who place their numbers on these lists share something in common, such as a greater desire for privacy, a representative sample of the general population cannot be obtained. Marketers and marketing researchers can obtain the do-not-call lists of phone numbers from the FTC for $62 per area code. The entire registry can be obtained for $17,050. This information can be obtained from the FTC do-not-call Web site shown in the screenshot on page 215. Although this may seem expensive, the FTC levies fines on the order of $10,000 per violation (per call) so obtaining the registry are a wise investment for those wishing to contact consumers via telephone. AT&T faced fines of over three-quarters of a million dollars for making 78 unwanted calls to 29 consumers listed on the do-not-call list. So, the Feds do take violations very seriously.

Likewise the Canadian government has instituted a nearly identical do-not-call program. The Canadian Radio-Television and Telecommunications Commission imposes fines up to $11,000 per call for calls made to people on the Canadian do-not-call list. Other countries in Europe and elsewhere are also considering such legislation. The advantages of privacy simply make phones less capable of obtaining representative samples than they once were. Often, however, a landline phone call is still the researcher's best option.

MOBILE PHONE INTERVIEWS

Mobile phone interviews differ from landline phones most obviously because they are directed toward a mobile (i.e., cell) phone number. However, there are other less obvious distinctions.

- In the U.S., no telemarketing can be directed toward mobile phone numbers by law. The primary reason for enacting this law was that respondents would often have to pay to receive the call. Respondents would have to "opt in" before their phone number would be made available for such calls.
- The recipient of a mobile phone call is even more likely to be distracted than with a call to someone's home or office. In fact, the respondent may be driving a car, on a subway train, or walking down a noisy street. Factors such as this are not conducive to a high quality interview.
- The area codes for mobile phones are not necessarily tied to geography. For instance, a person who moves from Georgia to Arizona can choose to keep his or her old phone number. Therefore, a researcher may be unable to determine whether or not a respondent fits into the desired geographic sampling population simply by taking note of the area code.
- The phones have varying abilities for automated responses and differing keypads. Some requests, such as "hit pound sign," may be more difficult to do on some keypads than on others.

Phone Interview Characteristics

Phone interviews in general have several distinctive characteristics that set them apart from other survey techniques. These characteristics present significant advantages and disadvantages for the researcher.

- **SPEED:** One advantage of telephone interviewing is the speed of data collection. While data collection with mail or personal interviews can take several weeks, hundreds of telephone interviews can be conducted literally overnight. When the interviewer enters the respondents' answers directly into a computerized system, the data processing speeds up even more.

- **COST:** As the cost of personal interviews continues to increase, telephone interviews are becoming relatively inexpensive. The cost of telephone interviews is estimated to be less than 25 percent of the cost of door-to-door personal interviews. Travel time and costs are eliminated. However, the typical Internet survey is less expensive than a telephone survey.

- **ABSENCE OF FACE TO FACE CONTACT:** Telephone interviews are more impersonal than face-to-face interviews. Respondents may answer embarrassing or confidential questions more willingly in a telephone interview than in a personal interview. However, mail and Internet surveys, although not perfect, are better media for gathering extremely sensitive information because they seem more anonymous. Some evidence suggests that people provide information on income and other financial matters only reluctantly, even in telephone interviews. Such questions may be personally threatening for a variety of reasons, and high refusal rates for this type of question occur with each form of survey research.

- **COOPERATION:** One trend is very clear. In the last few decades, telephone response rates have fallen. Analysis of response rates for the long-running Survey of Consumer Attitudes conducted by the University of Michigan found that response rates fell from a high of 72 percent to 67 percent during the period from 1979 to 1996 and then even faster after 1996, dropping to 60 percent.5 Lenny Murphy of data collection firm Dialtek says he has observed a decline in survey response rates from a typical range of 30 to 40 percent in the past down to below 20 percent.6 Fewer calls are answered because more households are using caller ID and answering machines to screen their calls, and many individuals do not pick up the phone when the display reads "out of area'' or when an unfamiliar survey organization's name and number appear on the display. Also, more phone lines are dedicated to fax machines and computers. However, the University of Michigan study found that the rate of refusal actually grew faster in the more recent period than the rate of not answering researchers' calls.

- **INCENTIVES TO REPOND:** Respondents should receive some incentive to respond. Research addresses different types of incentives. For telephone interviews, a test-marketing involving different type of survey introductions suggests that not all introductions are equally effective. A financial incentive or some significant chance to win a desirable prize will produce a higher telephone response rate than a simple assurance that the research is not a sales pitch, a more detailed description of the survey, or an assurance of confidentiality.

- **REPRESENTATIVE SAMPLES:** Practical difficulties complicate obtaining representative samples based on listings in the telephone book. About 95 percent of households in the United States have landline telephones. People without phones are more likely to be poor, aged, rural, or living in the South. Unlisted phone numbers and numbers too new to be printed in the directory are a greater problem. People have unlisted phone numbers for two reasons:
    - They have recently moved
    - They prefer to have unlisted numbers for privacy

  Individuals whose phone numbers are unlisted because of a recent move differ slightly from those with published numbers. The unlisted group tends to be younger, more urban, and less likely to own a single-family dwelling. Households that maintain unlisted phone numbers by choice tend to have higher incomes. And, as previously mentioned, a number of low-income households are unlisted by circumstance.

- **CALLBACKS:** An unanswered call, a busy signal, or a respondent who is not at home requires a callback. Telephone callbacks are much easier to make than callbacks in personal interviews. However, as mentioned, the ownership of telephone answering machines is growing, and their effects on callbacks need to be studied.

- **LIMITED DURATION:** Respondents who run out of patience with the interview can merely hang up. To encourage participation, interviews should be relatively short. The length of the

telephone interview is definitely limited.

- **LACK OF VISUAL MEDIUM:** Because visual aids cannot be used in telephone interviews, this method is not appropriate for packaging research, copy testing of television and print advertising, and concept tests that require visual materials. Likewise, certain attitude scales and measuring instruments, such as the semantic differential (described in a later chapter), require the respondent to see a graphic scale, so they are difficult to use over the phone.

Central Location Interviewing

Research agencies or interviewing services typically conduct all telephone interviews from a central location. Such **central location interviewing** allows firms to hire a staff of professional interviewers and to supervise and control the quality of interviewing more effectively. When telephone interviews are centralized and computerized, an agency or business can benefit from additional cost economies.

Computer-Assisted Telephone Interviewing

Advances in computer technology allow responses to telephone interviews to be entered directly into the computer in a process known as **computer-assisted telephone interviewing (CATI)**. Telephone interviewers are seated at computer terminals. Monitors display the questionnaires, one question at a time, along with precoded possible responses to each question. The interviewer reads each question as it appears on the screen. When the respondent answers, the interviewer enters the response directly into the computer, and it is automatically stored in the computer's memory. The computer then displays the next question on the screen. Computer-assisted telephone interviewing requires that answers to the questionnaire be highly structured. If a respondent gives an unacceptable answer (that is, one not precoded and programmed), the computer will reject it (see Research Snapshot above). Computer-assisted telephone interviewing systems include telephone management systems that select phone numbers, dial the numbers automatically, and perform other labor-saving functions. These systems can automatically control sample selection by randomly generating names or fulfilling a sample quota. A computer can generate an automatic callback schedule. A typical call management system might schedule recontact attempts to recall no answers after two hours and busy numbers after ten minutes and allow the interviewer to enter a more favorable time slot (day and hour) when a respondent indicates that he or she is too busy to be interviewed. Software systems also allow researchers to request daily status reports on the number of completed interviews relative to quotas. CATI interviews can also be conducted by a prerecorded voice with the respondent answering by punching buttons on the phone.

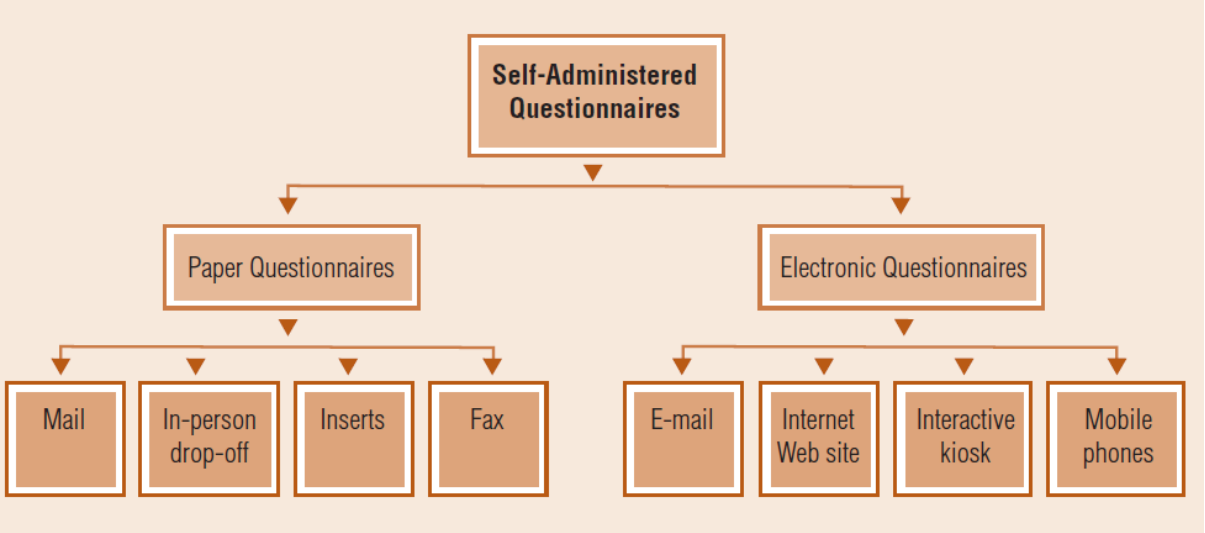
Computerized Voice-Activated Telephone Interview

Technological advances have combined computerized telephone dialing and voice-activated computer messages to allow researchers to conduct telephone interviews without human interviewers. However, researchers have found that computerized voice-activated telephone interviewing works best with very short, simple questionnaires. One system includes a voice- synthesized module controlled by a microprocessor. With it the sponsor is able to register a caller's single response such as "true/false," "yes/no," "like/dislike," or "for/against." This type of system has been used by television and radio stations to register callers' responses to certain issues. One system, Telsol, begins with an announcement that the respondent is listening to a recorded message. The computer then asks questions, leaving blank tape in between to record the answers. If respondents do not answer the first two questions, the computer disconnects and goes to the next call. With this process, the entire data collection process can be automated because a recorded voice is used to both ask the questions and record answers.

Self-Administered Questionnaires

Many surveys do not require an interviewer's presence. Researchers distribute questionnaires to consumers through the mail and in many other ways (see Exhibit 10.1). They insert questionnaires in packages and magazines. They may place questionnaires at points of purchase or in high-traffic locations in stores or malls. They may even fax questionnaires to individuals. Questionnaires can be printed on paper, but they may be posted on the Internet or sent via e-mail. No matter how the **self-administered questionnaires** are distributed, they are different from interviews because the respondent takes responsibility for reading and answering the questions.

f-Administered Questionnaires Can Be Either Printed or Electronic



Self-administered questionnaires present a challenge to the researcher because they rely on the clarity of the written word rather than on the skills of the interviewer. The nature of self- administered questionnaires is best illustrated by explaining mail questionnaires.

Mail Questionnaires

A **mail survey** is a self-administered questionnaire sent to respondents through the mail. This paper-and-pencil method has several **advantages and disadvantages**.

- **GEOGRAPHIC FLEXIBILITY:** Mail questionnaires can reach a geographically dispersed sample simultaneously because interviewers are not required. Respondents (such as farmers) who are located in isolated areas or those (such as executives) who are otherwise difficult to reach can easily be contacted by mail. For example, a pharmaceutical firm may find that doctors are not available for personal or telephone interviews. However, a mail survey can reach both rural and urban doctors who practice in widely dispersed geographic areas.
- **COST:** Mail questionnaires are relatively inexpensive compared with personal interviews, though they are not cheap. Most include follow-up mailings, which require additional postage and printing costs. And it usually isn't cost-effective to try to cut costs on printing—questionnaires photocopied on low-grade paper have a greater likelihood of being thrown in the wastebasket than those prepared with more expensive, high-quality printing. The low response rates contribute to the high cost.
- **RESPONDENT CONVENIENCE:** Mail surveys and other self-administered questionnaires can be filled out when the respondents have time, so respondents are more likely to take time to think about their replies. Many hard to reach respondents place a high value on convenience and thus are best contacted by mail. In some situations, particularly in business-to-business

research, mail questionnaires allow respondents to collect facts, such as employment statistics, that they may not be able to recall without checking. Being able to check information by verifying records or, in household surveys, by consulting with other family members should provide more valid, factual information than either personal or telephone interviews would allow. A catalog retailer may use mail surveys to estimate sales volume for catalog items by sending a mock catalog as part of the questionnaire. Respondents would be asked to indicate how likely they would be to order selected items. Using the mail allows respondents to consult other family members and to make their decisions within a reasonable time span.

- **ANONYMITY OF RESPONDENT:** In the cover letter that accompanies a mail or self-administered questionnaire, researchers almost always state that the respondents' answers will be confidential. Respondents are more likely to provide sensitive or embarrassing information when they can remain anonymous. For example, personal interviews and a mail survey conducted simultaneously asked the question "Have you borrowed money at a regular bank?" Researchers noted a 17 percent response rate for the personal interviews and a 42 percent response rate for the mail survey. Although random sampling error may have accounted for part of this difference, the results suggest that for research on personal and sensitive financial issues, mail surveys are more confidential than personal interviews. Anonymity can also reduce social desirability bias. People are more likely to agree with controversial issues, such as extreme political candidates, when completing self- administered questionnaires than when speaking to interviewers on the phone or at their doorsteps.
- ABSENCE OF INTERVIEWER: Although the absence of an interviewer can induce respondents to reveal sensitive or socially undesirable information, this lack of personal contact can also be a disadvantage. Once the respondent receives the questionnaire, the questioning process is beyond the researcher's control. Although the printed stimulus is the same, each respondent will attach a different personal meaning to each question. Selective perception operates in research as well as in advertising. The respondent does not have the opportunity to question the interviewer. Problems that might be clarified in a personal or telephone interview can remain misunderstandings in a mail survey. There is no interviewer to probe for additional information or clarification of an answer, and the recorded answers must be assumed to be complete. Respondents have the opportunity to read the entire questionnaire before they answer individual questions. Often the text of a later question will provide information that affects responses to earlier questions.
- **STANDARDIZED QUESTIONS:** Mail questionnaires typically are highly standardized, and the questions are quite structured. Questions and instructions must be clear-cut and straightforward. Ambiguous questions only create additional error. Interviewing allows for feedback from the interviewer regarding the respondent's comprehension of the questionnaire. An interviewer who notices that the first 50 respondents are having some difficulty understanding a question can report this fact to the research analyst so that revisions can be made. With a mail survey, however, once the questionnaires are mailed, it is difficult to change the format or the questions.
- **TIME IS MONEY:** If time is a factor in management's interest in the research results, or if attitudes are rapidly changing (for example, toward a political event), mail surveys may not be the best communication medium. A minimum of two or three weeks is necessary for receiving the majority of the responses. Follow-up mailings, which usually are sent when the returns begin to trickle in, require an additional two or three weeks. The time between the first mailing and the cut-off date (when questionnaires will no longer be accepted) normally is six to eight

weeks. In a regional or local study, personal interviews can be conducted more quickly. However, conducting a national study by mail might be substantially faster than conducting personal interviews across the nation.

- **LENGTH OF MAIL QUESTIONNAIRE:** Mail questionnaires vary considerably in length, ranging from extremely short postcard questionnaires to multipage booklets that require respondents to fill in thousands of answers. A general rule of thumb is that a mail questionnaire should not exceed six pages in length. When a questionnaire requires a respondent to expend a great deal of effort, an incentive is generally required to induce the respondent to return the questionnaire. The following sections discuss several ways to obtain high response rates even when questionnaires are longer than average.

Response Rates

All questionnaires that arrive via bulk mail are likely to get thrown away. Questionnaires that are boring, unclear, or too complex are even more likely to get thrown in the wastebasket. A poorly designed mail questionnaire may be returned by less than 5 percent of those sampled (that is, a 5 percent response rate). The basic calculation for obtaining a **response rate** is to count the number of questionnaires returned or completed, then divide the total by the number of eligible people who were contacted or requested to participate in the survey. Typically, the number in the denominator is adjusted for faulty addresses and similar problems that reduce the number of eligible participants.

The major limitations of mail questionnaires relate to response problems. Respondents who complete the questionnaire may not be typical of all people in the sample. Individuals with a special interest in the topic are more likely to respond to a mail survey than those who are indifferent.

A researcher has no assurance that the intended subject is the person who fills out the questionnaire. The wrong person answering the questions may be a problem when surveying corporate executives, physicians, and other professionals, who may pass questionnaires on to subordinates to complete. This probably is not unique to snail mail surveys since electronic surveying suffers similarly.

Evidence suggests that cooperation and response rates rise as home value increases. Also, if the sample has a high proportion of retired and well- off householders, response rates will be lower. Mail survey respondents tend to be better educated than nonrespondents. If they return the questionnaire at all, poorly educated respondents who cannot read and write well may skip open-ended questions to which they are required to write out their answers. Rarely will a mail survey have a 50 percent or greater response rate. However, the use of follow-up mailings and other techniques may increase the response rate to an acceptable percentage. The lower the response rate, the greater the concern that the resulting sample will not adequately represent the population.

Increasing Response Rates for Mail Surveys

Nonresponse error is always a potential problem with mail surveys. Individuals who are interested in the general subject of the survey are more likely to respond than those with less interest or little experience. Thus, people who hold extreme positions on an issue are more likely to respond than individuals who are largely indifferent to the topic. To minimize this bias, researchers have developed a number of techniques to increase the response rate to mail surveys. For example, almost all surveys include postage-paid return envelopes. Using a stamped return envelope instead of a business reply envelope increases response rates even more. Designing and formatting attractive questionnaires and wording questions so that they are easy to understand also help ensure a good response rate. However, special efforts may be required even with a sound questionnaire. Several of these methods are:

COVER LETTER

A **cover letter** that accompanies a questionnaire or is printed on the first page of the questionnaire booklet is an important means of inducing a reader to complete and return the questionnaire. Exhibit below illustrates a cover letter and some of the points considered by a research professional to be important in gaining respondents' attention and cooperation. The first paragraph of the letter explains why the study is important. The basic appeal alludes to the social usefulness of responding. Two other frequently used appeals are asking for help ("Will you do us a favor?") and the egotistical appeal ("Your opinions are important!"). Most cover letters promise confidentiality, invite the recipient to use an enclosed postage-paid reply envelope, describe any incentive or reward for participation, explain that answering the questionnaire will not be difficult and will take only a short time, and describe how the person was scientifically selected for participation.

A personalized letter addressed to a specific individual shows the respondent that he or she is important. Including an individually typed cover letter on letterhead rather than a printed form is an important element in increasing the response rate in mail surveys.11

MONEY HELPS

The respondent's motivation for returning a questionnaire may be increased by offering monetary incentives or premiums. Although pens, lottery tickets, and a variety of premiums have been used, monetary incentives appear to be the most effective and least biasing incentive. Money attracts attention and creates a sense of obligation. Perhaps for this reason, monetary incentives work for all income categories. Often, cover letters try to boost response rates with messages such as "We know that the attached dollar cannot compensate you for your time but please accept it as a token of our appreciation." Response rates increase dramatically when the monetary incentive is to be sent to a charity of the respondent's choice rather than directly to the respondent.

INTERESTING QUESTIONS

The topic of the research—and thus the point of the questions—cannot be manipulated without changing the definition of the research problem. However, certain interesting questions can be added to the questionnaire, perhaps at the beginning, to stimulate respondents' interest and to induce cooperation. By including questions that are of little concern to the researchers but that the respondents want to answer, the researchers may give respondents who are indifferent to the major questions a reason for responding.

FOLLOW UPS

Most mail surveys generate responses in a pattern like that shown in Exhibit on the next page, which graphs the cumulative response rates for two mail surveys. The response rates are relatively high for the first two weeks (as indicated by the steepness of each curve), then the rates gradually taper off. After responses from the first wave of mailings begin to trickle in, most studies use a follow up letter or postcard reminder, which request that the questionnaire be returned because a 100-percent return rate is important. A follow-up may include a duplicate questionnaire or may merely be a reminder to return the original questionnaire. Multiple contacts almost always increase response rates. The more attempts made to reach people, the greater the chances of their responding.

ADVANCE NOTIFICATION

Advance notification, by either letter or telephone, that a questionnaire will be arriving has been successful in increasing response rates in some situations. ACNielsen has used this technique to ensure a high cooperation rate in filling out diaries of television watching. Advance notices that go out closer to the questionnaire mailing time produce better results than those sent too far in advance. The optimal lead time for advance notification is three days before the mail survey is to arrive.

SURVEY SPONSORSHIP

Auspices bias may result from the sponsorship of a survey. One business-to-business researcher

wished to conduct a survey of its wholesalers to learn their stocking policies and their attitudes concerning competing manufacturers. A mail questionnaire sent on the corporate letterhead very likely would have received a much lower response rate than the questionnaire actually sent, which used the letterhead of a commercial marketing research firm. Sponsorship by well-known and prestigious organizations such as universities or government agencies may also significantly influence response rates. A mail survey sent to members of a consumer panel will receive an exceptionally high response rate because panel members have already agreed to cooperate with surveys.

OTHER TECHNIQUES

Numerous other devices have been used for increasing response rates. For example, the types of postage (commemorative versus regular stamp), envelope size, color of the questionnaire paper, and many other factors have been varied in efforts to increase response rates. Each has had at least limited success in certain situations; unfortunately, under other conditions each has failed to increase response rates significantly. The researcher should consider his or her particular situation. For example, the researcher who is investigating consumers faces one situation; the researcher who is surveying corporate executives faces quite another.

KEYING MAIL QUESTIONNAIRES WITHCODES

A researcher planning a follow-up letter or postcard should not disturb respondents who already have returned the questionnaire. The expense of mailing questionnaires to those who already have responded is usually avoidable. One device for eliminating those who have already responded from the follow-up mailing list is to mark the questionnaires so that they may be keyed to identify members of the sampling frame who are nonrespondents. Blind keying of questionnaires on a return envelope (systematically varying the job number or room number of the marketing research department, for example) or a visible code number on the questionnaire has been used for this purpose. Visible keying is indicated with statements such as “The sole purpose of the number on the last page is to avoid sending a second questionnaire to people who complete and return the first one.” Ethical researchers key questionnaires only to increase response rates, thereby preserving respondents’ anonymity.

Self-Administered Questionnaires Using Other Forms of Distribution

Many forms of self-administered, printed questionnaires are very similar to mail questionnaires. Airlines frequently pass out questionnaires to passengers during flights. Restaurants, hotels, and other service establishments print short questionnaires on cards so that customers can evaluate the service. *Tennis Magazine, Advertising Age, Wired,* and many other publications have used inserted questionnaires to survey current readers inexpensively, and often the results provide material for a magazine article. Many manufacturers use their warranty or owner registration cards to collect demographic information and data about where and why products were purchased. Using owner registration cards is an extremely economical technique for tracing trends in consumer habits. Again, problems may arise because people who fill out these self- administered questionnaires differ from those who do not. Extremely long questionnaires may be dropped off by an interviewer and then picked up later. The **drop-off method** sacrifices some cost savings because it requires traveling to each respondent’s location.

Fax Surveys

With fax surveys, potential survey respondents receive and/or return questionnaires via fax machines. A questionnaire inserted in a magazine may instruct the respondent to clip out the questionnaire and fax it to a certain phone number. In a mail survey, a prepaid-postage envelope places little burden on the respondent. But faxing a questionnaire to a long-distance number requires that the respondent pay for the transmission of the fax. Thus, a disadvantage of the **fax survey** is that only respondents with fax machines who are willing to exert the extra effort will return questionnaires. Again, people with extreme opinions will be more likely to respond. To address this disadvantage,

marketers may use faxing as one of several options for replying to a survey. Recently, the journal *American Family Physician* carried a reader survey that gave respondents the option of either returning the reply by fax or visiting the journal's Web site to answer the same questions online. For busy physicians who likely have access to office equipment, this approach would improve the response rate. Fax machines can also be used to distribute questionnaires. These fax surveys reduce the sender's printing and postage costs and can be delivered and returned faster than traditional mail surveys. Questionnaires distributed via fax can deal with timely issues. Although few households have fax machines, when the sample consists of organizations that are likely to have fax machines, the sample coverage may be adequate.

E-Mail Surveys

Questionnaires can be distributed via e-mail, but researchers must remember that some individuals cannot be reached this way. Certain projects do lend themselves to **e-mail surveys**, such as internal surveys of employees or satisfaction surveys of retail buyers who regularly deal with an organization via e-mail. The benefits of incorporating a questionnaire in an e-mail include the speed of distribution, lower distribution and processing costs, faster turnaround time, more flexibility, and less handling of paper questionnaires. The speed of e-mail distribution and the quick response time can be major advantages for surveys dealing with time-sensitive issues.

Not much academic research has been conducted on e-mail surveys. Nevertheless, some researchers have argued that many respondents feel they can be more candid in e-mail than in person or on the telephone, for the same reasons they are candid on other self-administered questionnaires. Yet in many organizations employees know that their e-mails are not secure and "eavesdropping" by a supervisor could possibly occur. Further, maintaining respondents' anonymity is difficult, because a reply to an e-mail message typically includes the sender's address. Researchers designing e-mail surveys should assure respondents that their answers will be confidential.

Not all e-mail systems have the same capacity: Some handle color and graphics well; others are limited to text. The extensive differences in the capabilities of respondents' computers and e-mail software limit the types of questions and the layout of the e-mail questionnaire. For example, the display settings for computer screens vary widely, and wrap- around of lines may put the questions and the answer choices into strange and difficult-to-read patterns.15 Many novice e-mail users find it difficult to mark answers in brackets on an e-mail questionnaire and/or to send a completed questionnaire using the e-mail reply function. For this reason, some researchers give respondents the option to print out the questionnaire, complete it in writing, and return it via regular mail. Unless the research is an internal organizational survey, this alternative, of course, requires the respondent to pay postage.

In general, the guidelines for printed mail surveys apply to e-mail surveys. However, some differences exist, because the cover letter and the questionnaire appear in a single e-mail message. A potential respondent who is not immediately motivated to respond, especially one who considers an unsolicited e-mail survey to be spam, can quickly hit the delete button to remove the e-mail. This response suggests that e- mail cover letters should be brief and the questionnaires relatively short. The cover letter should explain how the company got the recipient's name and should include a valid return e-mail address in the "from" box and reveal who is conducting the survey. Also, if the e-mail lists more than one address in the "to" or "CC" field, all recipients will see the entire list of names. This lack of anonymity has the potential to cause response bias and nonresponse error. When possible, the e-mail should be addressed to a single person. (The blind carbon copy, or BCC, field can be used if the same message must be sent to an entire sample.)

E-mail has another important role in survey research. E-mail letters can be used as cover letters asking respondents to participate in an Internet survey. Such e-mails typically provide a password and a link to a unique Web site location that requires a password for access.

Internet Surveys

An **Internet survey** is a self-administered questionnaire posted on a Web site. Respondents provide answers to questions displayed onscreen by highlighting a phrase, clicking an icon, or keying in an answer. Like every other type of survey, Internet surveys have both **advantages and disadvantages**.

SPEED AND COST EFFECTIVENESS

Internet surveys allow researchers to reach a large audience (possibly a global one), personalize individual messages, and secure confidential answers quickly and cost-effectively. These computer-to-computer self-administered questionnaires eliminate the costs of paper, postage, and data entry, as well as other administrative costs. Once an Internet questionnaire has been developed, the incremental cost of reaching additional respondents is minimal. So, samples can be larger than with interviews or other types of self-administered questionnaires. Even with large samples, surveys that used to take many weeks can be conducted in a week or less.

VISUAL APPEAL AND INTERACTIVITY

Surveys conducted on the Internet can be interactive. The researcher can use more sophisticated lines of questioning based on the respondents' prior answers. Many of these interactive surveys utilize color, sound, and animation, which may help to increase respondents' cooperation and willingness to spend time answering the questionnaires. The Internet is an excellent medium for the presentation of visual materials, such as photographs or drawings of product prototypes, advertisements, and movie trailers. Innovative measuring instruments that take advantage of the ability to adjust backgrounds, fonts, color, and other features have been designed and applied with considerable success.

RESPONDENT PARTICIPATION AND COOPERATION

Participation in some Internet surveys occurs because computer users intentionally navigate to a particular Web site where questions are displayed. For example, a survey of more than 10,000 visitors to the Ticketmaster Web site helped Ticketmaster better understand its customer purchase patterns and evaluate visitor satisfaction with the site. In some cases, individuals expect to encounter a survey at a Web site; in others, it is totally unexpected. In some instances, the visitor cannot venture beyond the survey page without providing information for the organization's "registration" questionnaire.

When the computer user does not expect a survey on a Web site and participation is voluntary, response rates are low. And, as with other questionnaires that rely on voluntary self-selection, participants tend to be more interested in or involved with the subject of the research than the average person. For many other Internet surveys, respondents are initially contacted via e-mail. Often they are members of consumer panels who have previously indicated their willingness to cooperate. When panel members receive an e-mail invitation to participate, they are given logon instructions and a password. This security feature prevents access by individuals who are not part of the scientifically selected sample. Assigning a unique password code also allows the researchers to track the responses of each respondent, thereby identifying any respondent who makes an effort to answer the questionnaire more than once. Panel members also need an incentive to respond.

A study of German consumers showed that nothing beat financial incentives. In other words, the best way to get responses was to simply pay consumers for participating in surveys. Ideally, the **welcome screen** contains the name of the research company and information about how to contact the organization if the respondent has a problem or concern. A typical statement might be "If you have any concerns or questions about this survey or if you experience any technical difficulties, please contact [name of research organization]."

REPRESENTATIVE SAMPLE

The population to be studied, the purpose of the research, and the sampling methods determine the quality of Internet samples, which varies substantially. If the sample consists merely of those who

visit a Web page and voluntarily fill out a questionnaire, then it is not likely to be representative of the entire U.S. population, because of self-selection error. However, if the purpose of the research is to evaluate how visitors feel about a Web site, randomly selecting every 100th visitor may accomplish the study's purpose. Scientifically drawn samples from a consumer panel, similar to what was done for the Digital Music Survey discussed above, or samples randomly generated in other ways also can be representative. Of course, a disadvantage, albeit ever decreasing, of Internet surveys is that many individuals in the general population cannot access the Internet. Even among people with Internet access, not all of them have the same level of technology.

Many people with low-speed Internet connections (low bandwidth) cannot quickly download high-resolution graphic files. Many lack powerful computers or software that is compatible with advanced features programmed into many Internet questionnaires. Some individuals have minimal computer skills. They may not know how to navigate through and provide answers to an Internet questionnaire. For example, the advanced audio- and video-streaming technology of RealPlayer or Windows Media Player software can be used to incorporate a television commercial and questions about its effectiveness into an Internet survey.

However, some respondents might find downloading the file too slow or even impossible, others might not have the RealPlayer or Windows Media Player software, and still others might not know how to use the streaming media software to view the commercial. For the foreseeable future, Internet surveys sampling the general public should be designed with the recognition that problems may arise for the reasons just described. Thus, photographs, animation, or other cutting-edge technological features created on the researcher's/Web designer's powerful computer may have to be simplified or eliminated so that all respondents can interact at the same level of technological sophistication. Because Internet surveys can be accessed anytime (24/7) from anywhere, they can reach certain hard-to-reach, busy respondents such as doctors, who would be almost impossible to reach via the telephone.

ACCURATEREAL TIME DATA CAPTURE

The computer-to-computer nature of Internet surveys means that each respondent's answers are entered directly into the researcher's computer as soon as the questionnaire is submitted. In addition, the questionnaire software may be programmed to reject improper data entry. For example, on a paper questionnaire a respondent might incorrectly check two responses even though the instructions call for a single answer. In an Internet survey, this mistake can be interactively corrected as the survey is taking place. Thus, the data capture is more accurate than when humans are involved. Real-time data capture allows for real-time data analysis. A researcher can review up-to-the minute sample size counts and tabulation data from an Internet survey in real time.

CALLBACKS

When the sample for an Internet survey is drawn from a consumer panel, those who have not completed the survey questionnaire can be easily recontacted. Computer software can simply automatically send e-mail reminders to panel members who did not visit the welcome page. Computer software can also identify the passwords of respondents who completed only a portion of the questionnaire and send those people customized messages. Sometimes such e-mails offer additional incentives to those individuals who terminated the questionnaire with only a few additional questions to answer, so that they are motivated to comply with the request to finish the questionnaire.

PERSONALIZED AND FLEXIBLEQUESTIONING

Computer-interactive Internet surveys are programmed in much the same way as computer assisted telephone interviews. That is, the software that is used allows questioning to branch off into two or more different lines depending on a respondent's answer to a filtered question. The difference is that there is no interviewer. The respondent interacts directly with software on a Web site. In other words,

the computer program asks questions in a sequence determined by the respondent's previous answers. The questions appear on the computer screen, and answers are recorded by simply pressing a key or clicking an icon, thus immediately entering the data into the computer's memory. Of course, these methods avoid labor costs associated with data collection and processing of paper-and-pencil questionnaires. This ability to sequence questions based on previous responses is a major advantage of computer-assisted surveys. The computer can be programmed to skip from question 6 to question 9 if the answer to question 6 is no. Furthermore, responses to previous questions can lead to questions that can be personalized for individual respondents (for example, "When you cannot buy your favorite brand, Revlon, what brand of lipstick do you prefer?").

Often the respondent's name appears in questions to personalize the questionnaire. Fewer and more relevant questions speed up the response process and increase the respondent's involvement with the survey. A related advantage of using a Web survey is that it can prompt respondents when they skip over a question. In a test comparing telephone and Internet versions of the same survey, the rate of item nonresponse was less for the Internet version, which issued a prompt for each item that was left blank. This was likely not a simple matter of motivation, because the rate of respondents who actually took the Web version was less than for the telephone version, even though the researchers offered a larger incentive to those who were asked to go online. (An earlier telephone screening had verified that everyone who was asked to participate had a computer.)

The ability to customize questions and the low cost per recipient also help researchers keep surveys short, an important consideration for boosting responses.18 Jakob Nielsen, a consultant on Internet usability with the Nielsen Norman Group, emphasizes that "quick and painless" surveys generate the highest response and urges researchers to keep surveys as short as possible. He suggests that if the research objectives call for a long survey, the questions can be divided among several questionnaires, with each version sent to a different group of respondents.

RESPONSE RATE

Many The methods for improving response rates for an Internet survey are similar to those for other kinds of survey research. A personalized invitation may be important. In many cases, the invitation is delivered via e-mail. The respondents may not recognize the sender's address, so the message's subject line is critical. The subject line should refer to a topic likely to interest the audience, and legal as well as ethical standards dictate that it may not be deceptive. Thus, the line might be worded in a way similar to the following: "Please give your opinion on [subject matter of interest]." Researchers should avoid gimmicks like dollar signs and the word *free,* either of which is likely to alert the spam filters installed on most computers.

As mentioned earlier, with a password system, people who have not participated in a survey in a predetermined period of time can be sent a friendly e-mail reminder asking them to participate before the study ends. This type of follow-up, along with preliminary notification, interesting early questions, and variations of most other techniques for increasing response rates to mail questionnaires, is recommended for Internet surveys. Unlike mail surveys, Internet surveys do not offer the opportunity to send a physical incentive, such as a dollar bill, to the respondent. Incentives to respond to a survey must be in the form of a promise of a future reward—for example, "As a token of appreciation for completing this survey, the sponsor of the survey will make a sizable contribution to a national charity. You can vote for your preferred charity at the end of the survey." Although some researchers have had success with promising incentives, academic research about Internet surveys is sparse, and currently there are few definitive answers about the most effective ways to increase response rates.

SECURITY CONCERNS

Many organizations worry that hackers or competitors may access Web sites to discover new product concepts, new advertising campaigns, and other top-secret ideas. Respondents may worry whether personal information will remain private. So may the organizations sponsoring the research.

Recently, McDonald's conducted quality-control research in England and Scotland, automating the transmittal of data with a system in which consultants used handheld devices and sent the numbers to headquarters as e-mail messages.

The system saved hours of work, but the company worried that confidential information could be compromised. McDonald's therefore purchased software that encrypted the data and allowed the handhelds to be remotely wiped clean of data if they were lost or stolen. As in the experience of McDonald's, no system can be 100 percent secure, but risks can be minimized. Many research service suppliers specializing in Internet surveying have developed password-protected systems that are very secure. One important feature of these systems restricts access and prevents individuals from filling out a questionnaire over and over again.

Kiosk Interactive Surveys

A computer with a touch screen may be installed in a kiosk at a trade show, at a professional conference, in an airport, or in any other high-traffic location to administer an interactive survey. Because the respondent chooses to interact with an on-site computer, self-selection often is a problem with this type of survey. Computer-literate individuals are most likely to complete these interactive questionnaires. At temporary locations such as conventions, these surveys often require a fieldworker to be at the location to explain how to use the computer system. This personal assistance is an obvious disadvantage.

Survey Research That Mixes Modes

For many surveys, research objectives dictate the use of some combination of telephone, mail, e-mail, Internet, and personal interview. For example, the researcher may conduct a short telephone screening interview to determine whether respondents are eligible for recontact in a more extensive personal interview. Such a **mixed-mode survey** combines the advantages of the telephone survey (such as fast screening) and those of the personal interview. A mixed-mode survey can employ any combination of two or more survey methods. Conducting a research study in two or more waves, however, creates the possibility that some respondents will no longer cooperate or will be unavailable in the second wave of the survey.

Several variations of survey research use cable television channels. For example, a telephone interviewer calls a cable subscriber and asks him or her to tune in to a particular channel at a certain time. An appointment is made to interview the respondent shortly after the program or visual material is displayed. NBC uses this type of mixed-mode survey to test the concepts for many proposed new programs.

Text-Message Surveys

Yes, surveys are even being sent via text messages. These may use the SMS (short message service) or MMS (multimedia message service). This technique is perhaps the newest survey approach. It has all the advantages of mobile phone surveys in terms of reach and it also shares the disadvantages in terms of reaching respondents who have not opted in via a mobile phone.

However, text-message surveys are catching on in other countries and are ideal for surveys involving only a few very short questions. Additionally, MMS messages can include graphic displays or even short videos. This technology is likely to see more applications in the near future.

Selecting the Appropriate Survey Research Design

Earlier discussions of research design and problem definition emphasized that many research tasks may lead to similar decision-making information. There is no best form of survey; each has advantages and disadvantages. A researcher who must ask highly confidential questions may use a mail survey, thus sacrificing speed of data collection to avoid interviewer bias. If a researcher must have

considerable control over question phrasing, central location telephone interviewing may be appropriate. To determine the appropriate technique, the researcher must ask several questions: Is the assistance of an interviewer necessary? Are respondents interested in the issues being investigated? Will cooperation be easily attained? How quickly is the information needed? Will the study require a long and complex questionnaire? How large is the budget? The criteria—cost, speed, anonymity, and so forth—may differ for each project. Exhibit on the next page summarizes the major advantages and disadvantages of typical door-to-door, mall intercept, telephone, mail, and Internet surveys. It emphasizes the typical types of surveys. For example, a creative researcher might be able to design highly versatile and flexible mail questionnaires, but most researchers use standardized questions. An elaborate mail survey may be far more expensive than a short personal interview, but generally this is not the case.

Pretesting

A researcher who is surveying 3,000 consumers does not want to find out after the questionnaires have been completed or returned that most respondents misunderstood a particular question, skipped a series of questions, or misinterpreted the instructions for filling out the questionnaire. To avoid problems such as these, screening procedures, or *pretests,* are often used. **Pretesting** involves a trial run with a group of respondents to iron out fundamental problems in the instructions or design of a questionnaire. The researcher looks for such obstacles as the point at which respondent fatigue sets in and whether there are any particular places in the questionnaire where respondents tend to terminate. Unfortunately, this stage of research is sometimes eliminated because of costs or time pressures. Broadly speaking, three basic ways to pretest exist. The first two involve screening the questionnaire with other research professionals, and the third—the one most often called pretesting— is a trial run with a group of respondents. When screening the questionnaire with other research professionals, the investigator asks them to look for such problems as difficulties with question wording, leading questions, and bias due to question order. An alternative type of screening might involve a client or the research manager who ordered the research. Often, managers ask researchers to collect information, but when they see the questionnaire, they find that it does not really meet their needs. Only by checking with the individual who has requested the questionnaire does the researcher know for sure that the information needed will be provided. Once the researcher has decided on the final questionnaire, data should be collected with a small number of respondents (perhaps 100) to determine whether the questionnaire needs refinement.

Ethical Issues in SurveyResearch

Many ethical issues apply to survey research, such as respondents' right to privacy, the use of deception, respondents' right to be informed about the purpose of the research, the need for confidentiality, the need for honesty in collecting data, and the need for objectivity in reporting data.

OBSERVATION METHODS

Observation in Business Research

In business research, **observation** is a systematic process of recording behavioral patterns of people, objects, and occurrences as they happen. No questioning or communicating with people is needed. Researchers who use observation as a method of data collection either witness and record information while watching events take place or take advantage of some tracking system such as check-out scanners or Internet activity records. These tracking systems can observe and provide data such as whether or not a specific consumer purchased more products on discount or at regular price or how long an employee takes to complete a specific task. Observation becomes a tool for scientific inquiry when it meets several conditions:

- The observation serves a formulated research purpose.
- The observation is planned systematically.
- The observation is recorded systematically and related to general propositions, rather than simply reflecting a set of interesting curiosities.
- The observation is subjected to checks or controls on validity and reliability

What Can Be Observed?

Observational studies gather a wide variety of information about behavior. Exhibit 11.1 lists seven kinds of observable phenomena*: physical actions,* such as shopping patterns (in-store or via a Web interface) or television viewing; *verbal behavior,* such as sales conversations or the exchange between a worker and supervisor; *expressive behavior,* such as tone of voice, facial expressions, or a coach stomping his foot; *spatial relations and locations,* such as traffic patterns; *temporal patterns,* such as amount of time spent shopping, driving, or making a business decision; *physical objects,* such as the amount of newspapers recycled or number of beer cans in the trash; and *verbal and pictorial records,* such as the content of advertisements or the number of minorities pictured in a company brochure.

Phenomenon	Example
Physical action	A worker's movement during an assembly process
Verbal behavior	Statements made by airline travelers while waiting in line
Expressive behavior	Facial expressions, tones of voices, and forms of body language
Spatial relations and locations	Proximity of middle managers' offices to the president's office
Temporal patterns	Length of time it takes to execute a stock purchase order
Physical objects	Percent of recycled materials compared to trash
Verbal and pictorial records	Number of illustrations appearing in a training booklet

While the observation method may be used to describe a wide variety of behavior, cognitive phenomena such as attitudes, motivations, and preferences cannot be observed. As a result, observation research cannot provide an explanation of why a behavior occurred or what actions were intended. Another limitation is that the observation period generally is short.

The Nature of Observation Studies

Business researchers can observe people, objects, events, or other phenomena using either human observers or machines designed for specific observation tasks. Human observation best suits a situation or behavior that is not easily predictable in advance of the research. Mechanical observation, as performed by supermarket scanners or traffic counters, can very accurately record situations or types of behavior that are routine, repetitive, or programmatic.

Human or mechanical observation is generally *unobtrusive;* meaning no communication with a respondent takes place. For example, rather than asking an employee how long it takes to handle an insurance claim, a researcher might observe and record the time it takes for different steps in this process. Or, rather than ask a consumer how long they spend shopping for produce, a researcher can watch shoppers in a supermarket and note the time each spends in the produce area. As noted in the opening vignette, the unobtrusive or nonreactive nature of the observation method often generates data without a subject's knowledge. A situation in which an observer's presence is known to the subject involves **visible observation**. A situation in which a subject is unaware that observation is taking place is **hidden observation**. Hidden, unobtrusive observation minimizes respondent error. Asking subjects to participate in the research is not required when they are unaware that they are being observed.

The major advantage of observation studies over surveys, which obtain self-reported data from respondents, is that the data are free from distortions, inaccuracies, or other response biases due to memory error, social desirability bias, and so on. The data are recorded when the actual behavior takes place.

Observation of HumanBehavior

Whereas surveys emphasize verbal responses, observation studies emphasize and allow for the systematic recording of nonverbal behavior. Toy manufacturers such as Fisher Price use the observation technique because children often cannot express their reactions to products. By observing children at play with a proposed toy, doll, or game, business researchers may be able to identify the elements of a potentially successful product. Toy researchers might observe play to answer the following questions:

- How long does the child's attention stay with the product?
- How exactly does the child play with the toy?
- Are the child's peers equally interested in the toy?

Behavioral scientists have recognized that nonverbal behavior can be a communication process by which meanings are exchanged among individuals. Head nods, smiles, raised eyebrows, and other facial expressions or body movements have been recognized as communication symbols. Observation of nonverbal communication may hold considerable promise for the business researcher. For example, a hypothesis about customer-salesperson interactions is that the salesperson would signal status based on the importance of each transaction. In low-importance transactions, in which potential customers are plentiful and easily replaced (say, a shoe store), the salesperson may show definite nonverbal signs of higher status than the customer. When customers are scarce, as in big-ticket purchase situations, the opposite should be true. For example, real estate sales agents may display nonverbal indicators of deference. One way to test this hypothesis would be with an observation study using the nonverbal communication measures shown in Exhibit below.

Behavior		Description	Example
Facial expressions		Expressions of emotion such as surprise (eyes wide open, mouth rounded and slightly open, brow furrowed)	A consumer reacts to the price quoted by a salesperson.
Body language		Posture, placement of arms and legs	A consumer crosses arms as salesperson speaks, possibly indicating a lack of trust.
Eye activity		Eye contact, staring, looking away, dilated pupils. In U.S. culture, not making eye contact is indicative of a deteriorating relationship. Dilated pupils can indicate emotion or degree of honesty.	A consumer avoids making eye contact with a salesperson knowing that he or she will not make a purchase.
Personal space		Physical distance between individuals; in the United States, people like to be about eight feet apart to have a discussion.	A consumer may back away from a salesperson who is viewed to be violating one's personal space.
Gestures		Responses to certain events with specific body reactions or gestures	A consumer who wins something (maybe at the casino or a sports contest) lifts arms, stands tall, and sticks out chest.
Manners		Accepted protocol for given situations	A salesperson may shake a customer's hand, but should not touch a customer otherwise.

Complementary Evidence

The results of observation studies may extend the results of other forms of research by providing *complementary evidence* concerning individuals' "true" feelings. Focus group interviews often are conducted behind two-way mirrors from which executives observe as well as listen to what is occurring. This additional source allows for interpretation of nonverbal behavior such as facial expressions or head nods to supplement information from interviews.

For example, in one focus group session concerning women's use of hand lotion, researchers observed that all the women's hands were above the table while they were casually waiting for the session to begin. Seconds after the women were told that the topic was to be hand lotion, all their hands were placed out of sight. This observation, combined with the group discussion, revealed the women's anger, guilt, and shame about the condition of their hands. Although they felt they were expected to have soft, pretty hands, their housework required them to wash dishes, clean floors, and do other chores that were hard on their hands. Note, however, that without the discussion provided by the participants the researcher would only have been able to note the action of placing their hands under the table, not the explanation for this behavior.

Direct Observation

Direct observation can produce detailed records of what people actually do during an event. The observer plays a passive role, making no attempt to control or manipulate a situation, instead merely recording what occurs. Many types of data can be obtained more accurately through direct observation than by questioning. For example, recording traffic counts or observing the direction of customer movement within a supermarket can help managers design store layouts that maximize the exposure of departments that sell impulse goods. A manufacturer can determine the number of facings, shelf locations, display maintenance, and other characteristics that improve store conditions. If directly questioned in a survey, most shoppers would be unable to accurately portray the time they spent in each department. The observation method, in contrast, could determine this without difficulty.

With the direct observation method, the data consist of records of events made as they occur. An observation form often helps keep researchers' observations consistent and ensures that they record all relevant information. A respondent is not required to recall—perhaps inaccurately—an event after it has occurred; instead, the observation is instantaneous.

In many cases, direct observation is the most straightforward form of data collection—or the only form possible. A produce manager for Auchan (a France-based hypermart firm) may periodically gather competitive price information from Carrefour (also a France-based hypermart firm) stores within competing areas. Both Carrefour and Auchan can monitor each other's promotions by observing promotions posted on the competitor's Web site (see **http://www.auchan.fr** and **http://www.carrefour.fr**, for example). In other situations, observation is the most economical technique. In a common type of observation study, a shopping center manager may observe the license plate (tag) numbers on cars in its parking lot. These data, along with automobile registration information, provide an inexpensive means of determining where customers live.

Certain data may be obtained more quickly or easily using direct observation than by other methods—gender, race, and other respondent characteristics can simply be observed. Researchers investigating a diet product may use observation when selecting respondents in a shopping mall. Overweight people may be prescreened by observing pedestrians, thus eliminating a number of screening interviews. Behaviors occurring in public places can also be easily observed, as the Research Snapshot on the next page shows.

In a quality-of-life survey, researchers asked respondents a series of questions that were compiled into an index of well-being. But interviewers also used direct observation because the researchers wanted to investigate the effect of weather conditions on people's answers. The researchers quickly and easily observed and recorded outside weather conditions on the day of the interviews, as well as the temperature and humidity in the building in which the interviews were conducted.

Recording the decision time necessary to make a choice between two alternatives is a relatively simple, unobtrusive task easily accomplished through direct observation. Observing the choice time as a measure of the strength of the preference between alternatives is called **response latency**. This measure is based on the hypothesis that the longer a decision maker takes to choose between two alternatives, the closer the two alternatives are in terms of preference. In contrast, making a quick decision presumably indicates a considerable psychological distance between alternatives—that is, the choice is obvious. It is simple for a computer to record decision times, so the response latency measure has gained popularity now that computer-assisted data collection methods are common.

ERRORS ASSOCIATED WITH DIRECT OBSERVATION

Although direct observation involves no interaction with the subject, the method is not error-free; the observer may record events subjectively. The same visual cues that may influence the interplay between interviewer and respondent (e.g., the subject's age or sex) may come into play in some direct observation settings, such as when the observer subjectively attributes a particular economic status or educational background to a subject. A distortion of measurement resulting from the cognitive behavior or actions of the witnessing observer is called **observer bias**. For example, in a research project using observers to evaluate whether sales clerks are rude or courteous, fieldworkers may be required to rely on their own interpretations of people or situations during the observation process.

Also, accuracy may suffer if the observer does not record every detail that describes the persons, objects, and events in a given situation. Generally, the observer should record as much detail as possible. However, the pace of events, the observer's memory and writing speed and other factors will limit the amount of detail that can be recorded.

Interpretation of observation data is another potential source of error. Facial expressions and other nonverbal communication may have several meanings. Does a smile always mean happiness?

Does the fact that someone is standing or seated next to the president of a company necessarily indicate the person's status?

SCIENTIFICALLY CONTRIVED OBSERVATION

Most observation takes place in a natural setting, but sometimes the investigator intervenes to create an artificial environment in order to test a hypothesis. This approach is called **contrived observation**. Contrived observation can increase the frequency of occurrence of certain behavior patterns, such as employee responses to complaints. An airline passenger complaining about a meal or service from the flight attendant may actually be a researcher recording that person's reactions. If situations were not contrived, the research time spent waiting and observing would expand considerably. This is one of the reasons for the growing popularity of *mystery shoppers* introduced in the opening vignette. They can effectively create a situation (such as a customer complaint) that might be very time consuming to observe if it were to occur naturally.

Combining Direct Observation and Interviewing

Some research studies combine visible observation with personal interviews. During or after detailed observations, individuals are asked to explain their actions.4 For example, direct observation of women applying hand and body lotion identified two kinds of users. Some women slapped on the lotion, rubbing it briskly into their skin. Others caressed their skin as they applied the lotion. When the women were questioned about their behavior, the researchers discovered that women who slapped the lotion on were using the lotion as a remedy for dry skin. Those who caressed their skin were more interested in making their skin smell nice and feel soft.

Ethical Issues in the Observation of Humans

Observation methods introduce a number of ethical issues. Hidden observation raises the issue of the respondent's right to privacy. Suppose a research firm is approached by a company interested in acquiring information about how women put on their bras. The researcher considers approaching spas in several key cities about placing small cameras inconspicuously to observe women getting dressed. Obviously, this is an illegal and unethical approach. However, what if other women are hired to observe and record this activity? While to some extent the dressing room is an area where women often do dress where others can observe them, women do not expect to have their dressing behavior recorded. Therefore, unless a way can be found to have some women consent to such observation, this observational approach is unethical.

Some people might see contrived observation as entrapment. To *entrap* means to deceive or trick into difficulty, which clearly is an abusive action. The problem is one of balancing values. If the researcher obtains permission to observe someone, the subject may not act naturally. So, at times there is a strong temptation to observe without obtaining consent. In other times, such as monitoring mall traffic, obtaining consent just to observe people walking through the mall would be difficult.

So, when should researchers feel comfortable collecting observational data? While exceptions exist to every rule, here are three questions that can help address this question:

1. Is the behavior being observed commonly performed in public where it is expected that others can observe the behavior?
2. Is the behavior performed in a setting in which the anonymity—meaning there is no way to identify individuals—of the person being observed is assured?
3. Has the person agreed to be observed?

If the answers to the first two questions are *yes*, then there is not likely a violation of privacy in collecting observational research data. If the answer to the third question is *yes*, then gathering the data also is likely to be ethical.

Observation of Physical Objects

Physical phenomena may be the subject of observation study. Physical-trace evidence is a visible mark of some past event or occurrence. For example, the wear on library books indirectly indicates which books are actually read (handled most) when checked out. A classic example of physical-trace evidence in a nonprofit setting was erosion on the floor tiles around the hatching- chick exhibit at Chicago's Museum of Science and Industry. These tiles had to be replaced every six weeks; tiles in other parts of the museum did not need to be replaced for years. The selective erosion of tiles, indexed by the replacement rate, was a measure of the relative popularity of exhibits.

Clearly, a creative business researcher has many options for determining the solution to a problem. The story about Charles Coolidge Parlin, generally recognized as one of the founders of commercial business research, counting garbage cans at the turn of the twentieth century illustrates another study of physical traces.

Parlin designed an observation study to persuade Campbell's Soup Company to advertise in the *Saturday Evening Post.* Campbell's was reluctant to advertise because it believed that the *Post* was read primarily by working people who would prefer to make soup from scratch, peeling the potatoes and scraping the carrots, rather than paying ten cents for a can of soup. To demonstrate that rich people weren't the target market, Parlin selected a sample of Philadelphia garbage routes. Garbage from each specific area of the city that was selected was dumped on the floor of a local National Guard Armory. Parlin had the number of Campbell's soup cans in each pile counted. The results indicated that the garbage from the rich people's homes didn't contain many cans of Campbell's soup. Although they may not have made soup from scratch themselves, their housekeepers may have. The garbage piles from the blue-collar area showed a larger number of Campbell's soup cans. This observation study was enough evidence for Campbell's. They advertised in the *Saturday Evening Post*.

The method used in this study has since been used in a scientific project at the University of Arizona in which aspiring archaeologists have sifted through garbage for over 30 years. They examine soggy cigarette butts, empty milk cartons, and half-eaten Big Macs in an effort to understand modern life.

What is most interesting about the garbage project is that observations can be compared with the results of surveys about food consumption—and garbage does not lie. This type of observation can correct for over-reporting consumption of healthful items and underreporting of, say, cigarette or alcohol consumption.

Another application of observing physical objects is to count and record physical inventories through retail or wholesale audits. This method allows researchers to investigate brand sales on regional and national levels, market shares, seasonal purchasing patterns, and so on. Business research suppliers offer audit data at both the retail and the wholesale levels.

An observer can record physical-trace data to discover information a respondent could not recall accurately. For example, measuring the number of ounces of a liquid bleach used during a test provides precise physical-trace evidence without relying on the respondent's memory. The accuracy of respondents' memories is not a problem for the firm that conducts a pantry audit. The pantry audit requires an inventory of the brands, quantities, and package sizes in a consumer's home rather than responses from individuals. The problem of untruthfulness or some other form of response bias is avoided. For example, the pantry audit prevents the possible problem of respondents erroneously claiming to have purchased prestige brands. However, gaining permission to physically check consumers' pantries is not easy, and the fieldwork is expensive. In addition, the brand in the pantry may not reflect the brand purchased most often if consumers substituted it because they had a coupon, the usual brand was out of stock, or some other reason.

Content Analysis

Besides observing people and physical objects, researchers may use **content analysis**, which obtains data by observing and analyzing the contents or messages of advertisements, newspaper articles, television programs, letters, and the like. This method involves systematic analysis as well as observation to identify the specific information content and other characteristics of the messages. Content analysis studies the message itself and involves the design of a systematic observation and recording procedure for quantitative description of the manifest content of communication.

This technique measures the extent of emphasis or omission of a given analytical category. For example, content analysis of advertisements might evaluate their use of words, themes, characters, or space and time relationships. Another topic of content analysis is the frequency with which women, African-Americans, or ethnic minorities appear in mass media.

Content analysis might be used to investigate questions such as whether some advertisers use certain themes, appeals, claims, or deceptive practices more than others or whether recent consumer-oriented actions by the Federal Trade Commission have influenced the contents of advertising. A cable television programmer might do a content analysis of network programming to evaluate its competition. Every year researchers analyze the Super Bowl telecast to see how much of the visual material is live-action play and how much is replay, or how many shots focus on the cheerleaders and how many on spectators. Content analysis also can explore the information content of television commercials directed at children, the company images portrayed in ads, and numerous other aspects of advertising.

Study of the content of communications is more sophisticated than simply counting the items; it requires a system of analysis to secure relevant data. After one employee role-playing session involving leaders and subordinates, researchers analyzed videotapes to identify categories of verbal behaviors (e.g., positive reward statements, positive comparison statements, and self-evaluation requests). Trained coders, using a set of specific instructions, then recorded and coded the leaders' behavior into specific verbal categories.

Mechanical Observation

In many situations, the primary—and sometimes the only—means of observation is mechanical rather than human. Video cameras, traffic counters, and other machines help observe and record behavior. Some unusual observation studies have used motion-picture cameras and time-lapse photography. An early application of this observation technique photographed train passengers and determined their levels of comfort by observing how they sat and moved in their seats. Another time-lapse study filmed traffic flows in an urban square and resulted in a redesign of the streets. Similar techniques may help managers determine how to better organize and arrange items in a warehouse or improve the design of store layouts to enhance traffic flow.

1-Television Monitoring

Perhaps the best-known research project involving mechanical observation and computerized data collection is ACNielsen's **television monitoring** system for estimating national television audiences. Nielsen Media Research uses a consumer panel and a monitoring device called a People Meter to obtain ratings for television programs nationwide. The Nielsen People Meter gathers data on what each television in a household is playing and who is watching it at the time. Researchers attach electronic boxes to television sets and remote controls to capture information on program choices and the length of viewing time. Nielsen matches the signals captured through these devices with its database of network broadcast and cable program schedules so that it can identify the specific programs being viewed.

When a television in the panel household is turned on, a red light on the People Meter periodically flashes to remind viewers to indicate who is watching. The viewer then uses a remote control to record who is watching. One button on the control is assigned to each member of the

household and a separate visitor button is used for potential guests. The household member presses his or her button to indicate the sex and age of the person who is watching. Knowing who in the family is watching allows executives to match television programs with demographic profiles. Each night, Nielsen's computers automatically retrieve the data stored in the People Meter's recording box. In this way, Nielsen gathers daily estimates of when televisions are in use, which channels are used, and who is viewing each program.

Other devices gather data about the viewing of advertisements. The TiVo digital television recorder, collects detailed viewing data, such as what commercials people skip by using fast-forward. The PreTesting Company sets up contrived observational studies in which viewers equipped with a remote control are invited to watch any of three prerecorded channels playing different programs and advertisements, including the client's ads to be tested. The system records the precise points at which the viewer changes the channel. By combining the results from many participants, the company arrives at a Cumulative Zapping Score, that is, the percentage of viewers who had exited the client's advertisement by each point in the ad. So that viewing behavior will be more natural, subjects are told they are evaluating the programming, not the ads.

2-Monitoring Web Site Traffic

Computer technology makes gathering detailed data about online behavior easy and inexpensive. The greater challenges are to identify which measures are meaningful and to interpret the data correctly. For instance, most organizations record the number of hits at their Web sites—mouse clicks on a single page of a Web site. If the visitor clicks on many links, that page receives multiple hits. Similarly, they can track *page views*, or single, discrete clicks to load individual pages of a Web site. Page views more conservatively indicate how many users visit each individual page on the Web site and may also be used to track the path or sequence of pages that each visitor follows:

CLICK THROUGH RATES

A **click-through rate** (CTR) is the percentage of people who are exposed to an advertisement who actually click on the corresponding hyperlink which takes them to the company's Web site. Counting hits or page views can suggest the amount of interest or attention a Web site is receiving, but these measures are flawed. First, hits do not differentiate between a lot of activity by a few visitors and a little activity by many visitors. In addition, the researcher lacks information about the meaning behind the numbers. If a user clicks on a site many times, is the person finding a lot of useful or enjoyable material, or is the user trying unsuccessfully to find something by looking in several places? Additionally, some hits are likely made by mistake. The consumers may have had no intention of clicking through the ad or may not have known what they were doing when they clicked on the ad.

A more refined count is the number of *unique visitors* to a Web site. This measurement counts the initial access to the site but not multiple hits on the site by the same visitor during the same day or week. Operators of Web sites can collect the data by attaching small files, called *cookies*, to the computers of visitors to their sites and then tracking those cookies to see whether the same visitors return. Some research companies, notably Jupiter Research and Nielsen//NetRatings, specialize in monitoring this type of Internet activity.

A typical approach is to install a special tracking program on the personal computers of a sample of Internet users who agree to participate in the research effort. Nielsen//NetRatings has its software installed in thirty thousand computers in homes and workplaces. Internet monitoring enables these companies to identify the popularity of Web sites. In recent years, accurate measurement of unique visitors has become more difficult, because over half of computer users have deleted cookies and many users block cookies to make themselves anonymous. As online advertising has become commonplace, business research has refined methods for measuring the effectiveness of the advertisements. The companies that place these ads can keep count of the click-through rate (CTR). Applying the CTR to the amount spent on the advertisement gives the advertiser a *cost per click*.

These measures have been hailed as a practical way to evaluate advertising effectiveness. However, marketers have to consider that getting consumers to click on an ad is rarely the ad's objective. Companies are more often advertising to meet short- or long-term sales goals.

Google has benefited from CTR research indicating that the highest click-through rates tend to occur on pages displaying search results. (Not surprisingly, someone who searches for the term *kayaks* is more likely to be interested in an advertisement offering a good deal on kayaks.) The company showed Vanguard, for example, that its banner ads cost the financial firm less than 50 cents per click and generated a 14 percent click-through rate. That CTR is far above typical response rates for direct-mail advertising, but it does not indicate whether online clicks are as valuable in terms of sales

Scanner-Based Research

Lasers performing optical character recognition and barcode technology like the universal product code (UPC) have accelerated the use of mechanical observation in business research. Chapter 8 noted that a number of syndicated services offer secondary data about product category movement generated from retail stores using scanner technology.

This technology allows researchers to investigate questions that are demographically or promotionally specific. Scanner research has investigated the different ways consumers respond to price promotions and the effects of those differences on a promotion's profitability. One of the primary means of implementing this type of research is through the establishment of a **scanner based consumer panel** to replace consumer purchase diaries. In a typical scanner panel, each household is assigned a bar-coded card, like a frequent-shopper card, which members present to the clerk at the register. The household's code number is coupled with the purchase information recorded by the scanner. In addition, as with other consumer panels, background information about the household obtained through answers to a battery of demographic and psychographic survey questions can also be coupled with the household code number.

Aggregate data, such as actual store sales as measured by scanners, are available to clients and industry groups. Data may also be aggregated by product category. To interpret the aggregated data, researchers can combine them with secondary research and panel demographics. For instance, data from Information Resources Inc. (IRI) have indicated a downward trend in sales of hair-coloring products. Demographic data suggest that an important reason is the aging of the population; many consumers who dye their hair reach an age at which they no longer wish to cover their gray hair. A smaller segment of the population is at an age where consumers typically begin using hair coloring.

Data from scanner research parallel data provided by a standard mail diary panel, with some important improvements:

1. The data measure observed (actual) purchase behavior rather than reported behavior (recorded later in a diary).
2. Substituting mechanical for human record-keeping improves accuracy.
3. Measures are unobtrusive, eliminating interviewing and the possibility of social desirability or other bias on the part of respondents.
4. More extensive purchase data can be collected, because all UPC categories are measured. In a mail diary, respondents could not possibly reliably record all items they purchased. Because all UPC-coded items are measured in the panel, users can investigate many product categories to determine loyalty, switching rates, and so on for their own brands as well as for other companies' products and locate product categories for possible market entry.
5. The data collected from computerized checkout scanners can be combined with data about the timing of advertising, price changes, displays, and special sales promotions. Researchers can scrutinize them with powerful analytical software provided by the scanner data providers.

Scanner data can show a researcher week-by-week how a product is doing, even in a single store, and track sales in response to changes of sales personnel, local advertising, or price promotions.

Also, several organizations have developed scanner panels, such as Information Resources Inc.'s Behavior Scan System, and expanded them into electronic test-market systems.

Measuring Physiological Reactions

Researchers have used a number of mechanical devices to evaluate physical and physiological reactions to advertising copy, packaging, and other stimuli. Researchers use such means when they believe consumers are unaware of their own reactions to stimuli such as advertising or that consumers will not provide honest responses. Recent research approaches use devices to monitor and measure brain activity as described in the Research Snapshot above. Four major categories of mechanical devices are used to measure physiological reactions:

(1) Eye-Tracking Monitors,
(2) Pupilometers,
(3) Psychogalvanometers, And
(4) Voice-Pitch Analyzers.

A magazine or newspaper advertiser may wish to grab readers' attention with a visual scene and then direct it to a package or coupon. Or a television advertiser may wish to identify which selling points to emphasize. Eye-tracking equipment records how the subject reads a print ad or views a television commercial and how much time is spent looking at various parts of the stimulus. In physiological terms, the gaze movement of a viewer's eye is measured with an **eye-tracking monitor**, which measures unconscious eye movements. Originally developed to measure astronauts' eye fatigue, modern eye-tracking systems need not keep a viewer's head in a stationary position. The devices track eye movements with invisible infrared light beams that lock onto a subject's eyes. The light reflects off the eye, and eye-movement data are recorded while another tiny video camera monitors which magazine page is being perused.

The data are analyzed by computer to determine which components in an ad (or other stimuli) were seen and which were overlooked. Eye-tracking monitors have recently been used to measure the way subjects view e-mail and Web messages. OgilvyOne has used this technology to learn that people often skip over more than half of the words in e-mail advertising, especially words on the right side of the message Interestingly, consumers generally ignore the word *free*. Other physiological observation techniques are based on a common principle: that adrenaline is released when the body is aroused. This hormone causes the heart to enlarge and to beat harder and faster. These changes increase the flow of blood to the fingers and toes. The blood vessels dilate, and perspiration increases, affecting the skin's electrical conductivity. Other physical changes following the release of adrenaline include dilation of the pupils, more frequent brain wave activity, higher skin temperature, and faster breathing. Methods that measure these and other changes associated with arousal can apply to a variety of business questions, such as subjects' reactions to advertising messages or product concepts.

A **pupilometer** observes and records changes in the diameter of a subject's pupils. A subject is instructed to look at a screen on which an advertisement or other stimulus is projected. When the brightness and distance of the stimulus from the subject's eyes are held constant, changes in pupil size may be interpreted as changes in cognitive activity that result from the stimulus, rather than from eye dilation and constriction in response to light intensity, distance from the object, or other physiological reactions to the conditions of observation. This method of research is based on the assumption that increased pupil size reflects positive attitudes toward and interest in advertisements.

A **psychogalvanometer** measures galvanic skin response (GSR), a measure of involuntary changes in the electrical resistance of the skin. This device is based on the assumption that physiological changes, such as increased perspiration, accompany emotional reactions to advertisements, packages, and slogans. Excitement increases the body's perspiration rate, which increases the electrical resistance of the skin. The test is an indicator of emotional arousal or tension.

Voice- pitch analysis is a relatively new physiological measurement technique that gauges emotional reactions as reflected in physiological changes in a person's voice. Abnormal frequencies in

the voice caused by changes in the autonomic nervous system are measured with sophisticated, audio-adapted computer equipment. Computerized analysis compares the respondent's voice pitch during warm-up conversations (normal range) with verbal responses to questions about his or her evaluative reaction to television commercials or other stimuli. This technique, unlike other physiological devices, does not require the researcher to surround subjects with mazes of wires or equipment. All of these devices assume that physiological reactions are associated with persuasiveness or predict some cognitive response. This assumption has not yet been clearly demonstrated. No strong theoretical evidence supports the argument that such a physiological change is a valid measure of future sales, attitude change, or emotional response.

Another major problem with physiological research is the *calibration,* or sensitivity, of measuring devices. Identifying arousal is one thing, but precisely measuring *levels* of arousal is another. In addition, most of these devices are expensive. However, as a prominent researcher points out, physiological measurement is coincidental: "Physiological measurement isn't an exit interview. It's not dependent on what was remembered later on. It's a live blood, sweat, and tears, moment- by-moment response, synchronous with the stimulus." Each of these mechanical devices has another limitation: The subjects are usually placed in artificial settings, such as watching television in a laboratory rather than at home, and the participants know they are being observed.

EXPERIMENTAL RESEARCH

Creating an Experiment

As described in an earlier chapter, experiments are widely used in causal research designs. Experimental research allows a researcher to control the research situation so that *causal* relationships among variables may be evaluated. The experimenter manipulates one or more independent variables and holds constant all other possible independent variables while observing effects on dependent variable(s). Events may be controlled in an experiment to a degree that is simply not possible in a survey.

Independent variables are expected to determine the outcomes of interest. In an experiment, they are controlled by the researcher through manipulations. Dependent variables are the outcomes of interest to the researcher and the decision makers. A simple example would be thinking about how changes in price would influence sales. Price would be an independent variable and sales would be a dependent variable. In our opening vignette, the protocol used (FIX or SOAP) would be an experimental manipulation—the independent variable—and the speed of data transmission is the important dependent variable.

The researcher's goal in conducting an experiment is to determine whether changing an experimental independent variable causes changes in the specified dependent variable. The assumption of the experiment described above is that the type of protocol used will affect the speed of financial data transfer. In other words, changing from FIX to SOAP will increase or decrease the time required for data transfer. If all the other conditions are the same, then a causal inference is supported.

A famous experiment in the marketing field investigated the influence of brand name on consumers' taste perceptions. An experimenter manipulated whether consumers preferred the taste of beer in labeled or unlabeled bottles. One week respondents were given a six-pack containing bottles labeled only with letters (A, B, C). The following week, respondents received another six-pack with brand labels (like Budweiser, Coors, Miller, and so forth). The experimenter measured reactions to the beers after each tasting. In every case, the beer itself was the same. So, every person involved in the experiment drank the very same beer. Therefore, the differences observed in taste, the key dependent variable, could only be attributable to the difference in labeling. When the consumers participating in the experiment expressed a preference for the branded beer, the conclusion is that brand name does influence consumers' taste perceptions.

An Illustration: Can a Self-Efficacy Intervention Enhance Job Attitude?

Self-efficacy is a person's confidence and belief in their own abilities to accomplish the tasks at hand. While the subjects of this particular research are accountants, and the results are highly relevant for those involved in fields like human resource management, it has implications for anyone in a managerial role. The key issue centers on a manager's ability to raise an employee's confidence in their ability to perform their job and the favorable outcomes of this increased confidence.

EXPERIMENTAL SUBJECTS

This experiment involved actual employees of an accounting firm. Seventy-one first and second year auditors of one major accounting firm participated in the study. Participants in experimental research are referred to as **subjects** rather than respondents. This is because the experimenter subjects them to some experimental treatment. In this experiment, 35 of the subjects were given positive feedback and encouragement from their supervisors as the experimental treatment. The other 36 subjects were not provided the positive feedback.

INDEPENDENT VARIABLES

The experiment involved one relevant independent variable, whether or not the employee received the positive feedback intended to enhance their self-efficacy. Employees receiving the experimental treatment participated in an interview and received three different pieces of written communication from their supervisors providing encouragement and expressing confidence that they would be successful in their positions.

While not a true independent variable, the length of time each employee had worked with the firm was also important to the researchers. Could new employees react differently to the positive feedback than employees who had already been with the firm? In this case, length of time cannot be manipulated by the researchers, but it can still be considered in the experiment. Variables such as this (another example would be the sex of the experimental subject) are referred to as **blocking variables**, which are discussed in more detail later in this chapter. Considering the independent variable (treatment or no treatment) and the blocking variable (new or current employee), four different experimental cells are possible. Exhibit below illustrates the four different experimental conditions for this experiment. An **experimental condition** refers to one of the possible levels of an experimental variable manipulation.

Subjects were divided into "newcomers" (the new employees) and "insiders" (current employees) and then randomly assigned to either the treatment condition or the control group. By analyzing differences between the groups, the researcher can see what effects occur due to the independent and blocking variables.

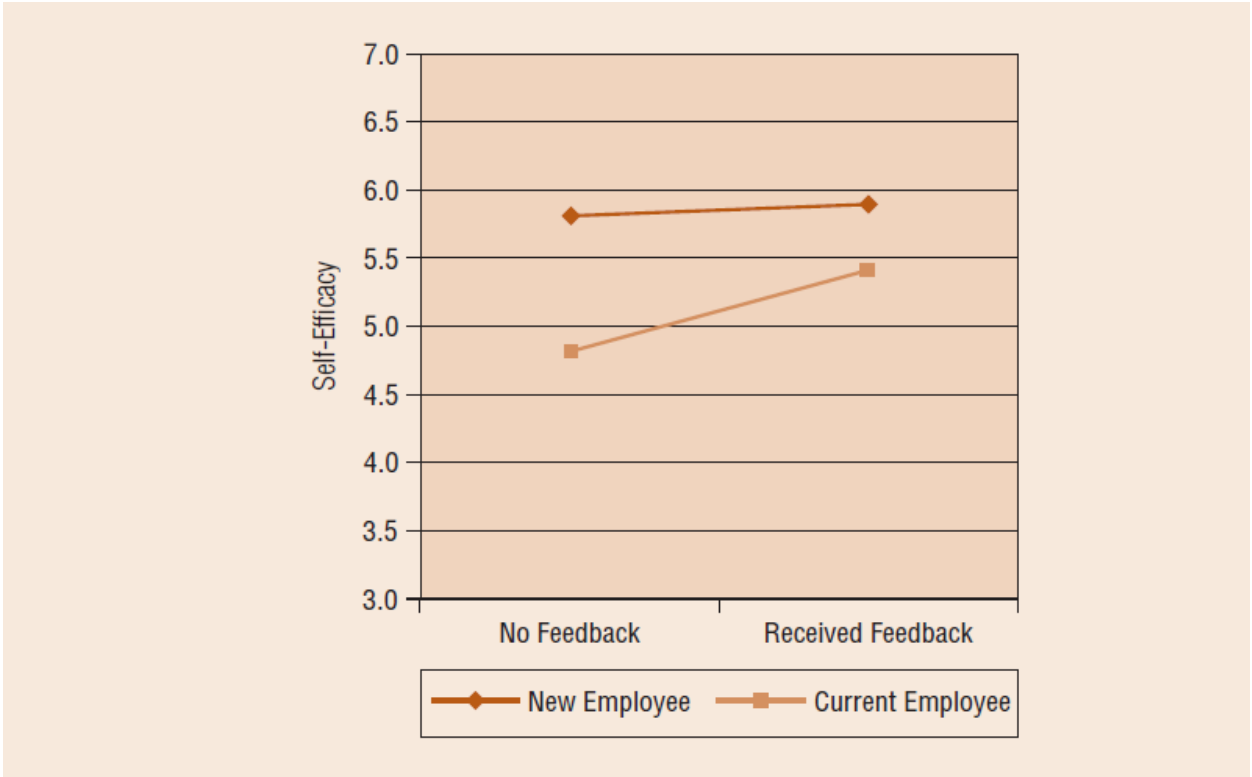
EXPERIMENTAL OUTCOME

The key outcomes, or dependent variables, in this example are a subject's job satisfaction, organizational commitment, professional commitment, intent to quit the organization, and intent to quit the profession. In addition, the researchers followed up later to see if the subjects had actually left the firm. For simplicity, we will only look at the effect on one dependent variable, job satisfaction. In this case, subjects were asked to respond to a rating scale asking how much they agreed with a series of statements regarding their satisfaction with their job. The possible scores ranged from 1 to 7, where a higher score means higher job satisfaction. Exhibit below shows the average for each experimental condition. The results show that after receiving the positive feedback and encouragement, subjects that were already working for the firm reported an average job satisfaction score of 5.45, while the new employees that received the treatment reported an average score of 5.93. For those subjects that did not receive the treatment, the current employees' average scores was 4.77 and the new employee's average was 5.80.

INDEPENDENT VARIABLE MAIN EFFECTS AND INTERACTION

The length of time that the employee works at the firm clearly appears to matter. But maybe the attempts to enhance self-efficacy shouldn't be dismissed so quickly. The researcher must examine both the effects of each experimental variable considered alone and the effects due to combinations of variables. A **main effect** refers to the experimental difference in means between the different levels of any single experimental variable. In this case, there are potential main effects for the self-efficacy treatment and for the length of time as an employee, but only the differences associated with employment length are significant (at a .05 level).

An **interaction effect** is due to a specific combination of independent variables. In this case, it's possible that the combination of length of employment and the self-efficacy treatment creates effects that are not clearly represented in the main effects. Interaction results are often shown with a line graph as shown in Exhibit below. Main effects are illustrated when the lines are at different heights as is the case here. Notice the line for new employees is higher than the line for current employees. When the lines have very different slopes, an interaction is likely present. In this case, the combination of length of employment and self-efficacy treatment is presenting an interaction leading to the following interpretation.



The worst situation is the current employees who do not receive positive feedback. Conversely, the best scenario regarding job satisfaction occurs when the treatment is given to new employees. It also appears that job satisfaction tends to decrease over time. The benefit of the self-efficacy treatment is greater for the employees that have been with the organization than for new employees. In other words, it appears that the self-efficacy treatment helps prevent the decline in job satisfaction.

Designing an Experiment to Minimize Experimental Error

Experimental design is a major research topic. In fact, there are courses and books devoted only to that topic.3 Here, an introduction into experimental design is provided. A student should be able to design and implement basic experimental designs with this introduction. Fortunately, most experimental designs for business research are relatively simple. Experimental designs involve no less than four

important design elements. These issues include:

(1) manipulation of the independent variable(s);
(2) selection and measurement of the dependent variable(s);
(3) selection and assignment of experimental subjects; and
(4) control over extraneous variables.

Each element can be implemented in a way that helps minimize error.

Manipulation of the Independent Variable

The thing that makes independent variables special in experimentation is that the researcher actually creates his or her values. This is how the researcher manipulates, and therefore controls, independent variables. In the financial market protocol experiment, the researchers decided to test different protocols. In the self-efficacy example, the researchers chose to provide some employees with positive feedback and not give the same encouragement to others. Experimental independent variables are hypothesized to be causal influences. Therefore, experiments are very appropriate in causal designs.

An **experimental treatment** is the term referring to the way an experimental variable is manipulated. For example, the opening vignette manipulated the protocol by choosing FIX and SOAP for their test. In the self efficacy study, the researchers had a personal interview with the employees and then had the supervisors send three encouraging letters to manipulate self efficacy. Similarly, a medical researcher may manipulate an experimental variable by treating some subjects with one drug and the other subjects with a separate drug. Experimental variables often involve treatments with more than two levels. For instance, prices of $1.29, $1.69, and $1.99 might represent treatments in a pricing experiment examining how price affects sales.

Experimental variables like these can not only be described as independent variables, but they also can be described as a *categorical variable* because they take on a value to represent some classifiable or qualitative aspect. Protocol, for example, is either FIX or SOAP. The employees either received the feedback or they did not. In other situations an independent variable may truly be a *continuous variable.* For example, the pricing experiment mentioned above could involve any price levels. The task for the researcher is to select appropriate levels of that variable as experimental treatments. For example, consumers might not perceive a difference between $1.24 and $1.29, but likely will notice the difference between $1.29 and $1.69. Before conducting the experiment, the researcher decides on levels that would be relevant to study. The levels should be noticeably different and realistic.

EXPERIMENTAL AND CONTROL GROUPS

In perhaps the simplest experiment, an independent variable is manipulated over two treatment levels resulting in two groups, an experimental group and a control group. An **experimental group** is one in which an experimental treatment is administered. A **control group** is one in which no experimental treatment is administered. In our self-efficacy example, the experimental group is comprised of the subjects that received the positive feedback. The control group did not receive the additional positive feedback designed to enhance self-efficacy. By holding conditions constant in the control group, the researcher controls for potential sources of error in the experiment. Job satisfaction (the dependent variable) in the two groups was compared at the end of the experiment to determine whether the encouragement (the independent variable) had any effect.

SEVERAL EXPERIMENTAL TREATMENT LEVELS

An experiment with one experimental and one control group may not tell a manager everything he or she wishes to know. If an advertiser wished to understand the functional nature of the

relationship between advertising and sales at several treatment levels, additional experimental groups with annual advertising expenditures of $250,000, $500,000, $750,000, and $1 million might be studied. This experiment may still involve a control variable (keeping the advertising budget for a region at the current level of $100,000). By analyzing more groups each with a different treatment level, a more precise result may be obtained than in a simple experimental group–control group experiment. This design, only manipulating the level of advertising, can produce only a main effect.

MORE THAN ONE INDEPENDENTVARIABLE

An experiment can also be made more complicated by including the effect of another experimental variable. Our extended example of the self-efficacy experiment would typify a still relatively simple two-variable experiment. Since there are two variables, each with two different levels, four experimental groups are obtained. Often, the term **cell** is used to refer to a treatment combination within an experiment. The number of cells involved in any experiment can be easily computed as follows:

$$K = (T1)(T2)...(Tm)$$

where K = the number of cells, $T1$ = the number of treatment levels for experimental group number one, $T2$ = the number of treatment levels for experimental group number two, and so forth through the m^{th} experimental group (Tm). In this case, since there are two variables each with two levels, the computation is quite simple:

$$K = 2 \times 2 = 4 \; cells$$

Including multiple variables allows a comparison of experimental treatments on the dependent variable. Since there are more than two experimental variables, this design involves both main effects and interactions.

REPEATED MEASURES

Experiments in which an individual subject is exposed to more than one level of an experimental treatment are referred to as **repeated measures** designs. Although this approach has advantages, including being more economical since the same subject provides more data than otherwise, it has several drawbacks that can limit its usefulness.

Selection and Measurement of the Dependent Variable

Selecting dependent variables is crucial in experimental design. Unless the dependent variables are relevant and truly represent an outcome of interest, the experiment will not be useful. Sometimes, the dependent variable is fairly obvious. In the protocol example, the speed of the data exchange is an important and logical dependent measure. Other dependent measures, such as number of errors in the data transmission, may also be of relevance. In the self-efficacy study, the researchers did consider several dependent measures in addition to job satisfaction (organizational commitment, professional commitment, intent to quit the profession, intent to quit the organization, and actual turnover). In some situations, however, clearly defining the dependent variable is not so easy. If researchers are experimenting with different forms of advertising copy appeals, defining the dependent variable may be more difficult. For example, measures of advertising awareness, recall, changes in brand preference, or sales might be possible dependent variables.

Choosing the right dependent variable is part of the problem definition process. Like the problem definition process in general, it sometimes is considered less carefully than it should be. The experimenter’s choice of a dependent variable determines what type of answer is given to assist managers in decision making.

Consider how difficult one might find selecting the right dependent variable in an advertising experiment. While sales are almost certainly important, when should sales be measured? What about brand image or recognition? The amount of time needed for effects to become evident should be considered in choosing the dependent variable. Sales may be measured several months after the changes in advertising to determine if there were any carryover effects. Changes that are relatively permanent or longer lasting than changes generated only during the period of the experiment should be considered. Repeat purchase behavior may be important too, since the advertising may motivate some consumers to try a product once, but then never choose that product again. Consumers often try a "loser" once, but they do not buy a "loser" again and again.

The introduction of the original Crystal Pepsi illustrates the need to think beyond consumers' initial reactions. When Crystal Pepsi, a clear cola, was introduced, the initial trial rate was high, but only a small percentage of customers made repeat purchases. The brand never achieved high repeat sales within a sufficiently large market segment. Brand awareness, trial purchase, and repeat purchase are all possible dependent variables in an experiment. The dependent variable therefore should be considered carefully. Thorough problem definition will help the researcher select the most important dependent variable or variables.

Selection and Assignment of Test Units

Test units are the subjects or entities whose responses to the experimental treatment are measured or observed. Individual consumers, employees, organizational units, sales territories, market segments, or other entities may be the test units. People, whether as customers or employees, are the most common test units in most organizational behavior, human resources, and marketing experiments.

SAMPLE SELECTION AND RANDOM SAMPLING ERRORS

As in other forms of business research, random sampling errors and sample selection errors may occur in experimentation. For example, experiments sometimes go awry even when a geographic area is specially chosen for a particular investigation. A case in point was the experimental testing of a new lubricant for outboard boat motors by Dow Chemical Company. The lubricant was tested in Florida. Florida was chosen because researchers thought the hot, muggy climate would provide the most demanding test. In Florida the lubricant was a success. However, the story was quite different when the product was sold in Michigan. Although the lubricant sold well and worked well during the summer, the following spring Dow discovered the oil had congealed, allowing the outboard motors, idle all winter, to rust. The rusting problem never came to light in Florida, where the motors were in year-round use. Thus, sample selection error occurs because of flaws in procedures used to assign experimental test units. Florida conditions made the experiment irrelevant in Michigan.

Systematic or nonsampling error may occur if the sampling units in an experimental cell are somehow different than the units in another cell, and this difference affects the dependent variable. For example, suppose some professors are interested in testing the effect of providing snacks during exams on student's scores. The experimental variable is snacks, manipulated over three levels: (1) fruit, (2) cookies, and (3) chocolate. The test units in this case are individual students. When the professors conduct the experiment, for convenience, they decide to give all of the 8:00 a.m. classes chocolate for a snack, all of the 1:00 p.m. classes get fruit, and all of the 7:00 p.m. classes get cookies. While this type of procedure is often followed, if our tastes and digestive systems react differently to different foods at different times of the day, systematic error is introduced into the experiment. Furthermore, because the night classes contain students who are older on average, the professors may reach the conclusion that students perform better when they eat cookies, when it may really be due to the fact that students who are older perform better no matter what they are fed.

RANDOMIZATION

Randomization—the random assignment of subject and treatments to groups—is one device for equally distributing the effects of extraneous variables to all conditions. These **nuisance variables**, items that may affect the dependent measure but are not of primary interest, often cannot be eliminated. However, they will be controlled because they are likely to exist to the same degree in every experimental cell if subjects are randomly assigned. In our self-efficacy experiment, it is likely that some subjects are happier with their positions to start with, have greater or lesser ability, and so forth. By randomly assigning employees to the control and experimental group, all these factors should balance out. Thus, all cells would be expected to yield similar average scores on the dependent variables if it were not for the experimental treatment administered. In other words, the researcher would like to set up a situation where everything in every cell is the same except for the experimental treatment. Random assignment of subjects allows the researcher to make this assumption.

MATCHING

Random assignment of subjects to the various experimental groups is the most common technique used to prevent test units from differing from each other on key variables; it assumes that all characteristics of the subjects have been likewise randomized. Matching the respondents on the basis of pertinent background information is another technique for controlling systematic error by assigning subjects in a way that their characteristics are the same in each group. This is best thought of in terms of demographic characteristics. If a subject's sex is expected to influence dependent variable responses, as in a taste test, then the researcher may make sure that there are equal numbers of men and women in each experimental cell. In general, if a researcher believes that certain extraneous variables may affect the dependent variable, he or she can make sure that the subjects in each group are the same on these characteristics.

CONTROL OVER EXTRANEOUSVARIABLES

The fourth decision about the basic elements of an experiment concerns control over extraneous variables. This is related to the various types of experimental error. In an earlier chapter, we classified total survey error into two basic categories: random sampling error and systematic error. The same dichotomy applies to all research designs, but the terms *random (sampling) error* and *systematic error* are more frequently used when discussing experiments.

EXPERIMENTAL CONFOUNDS

We have already discussed how systematic error can occur when the extraneous variables or the conditions of administering the experiment are allowed to influence the dependent variables. When this occurs, the results will be confounded because the extraneous variables have not been controlled or eliminated. A **confound** means that there is an alternative explanation beyond the experimental variables for any observed differences in the dependent variable. Once a potential confound is identified, the validity of the experiment is severely questioned.

Recall from the opening vignette the experimental procedures involved in the protocol test. The same data was sent over the two protocols in the experiment. What if the FIX protocol was better suited to handling small files, while the SOAP protocol was better suited for large files? If only small files were tested, the experiment has a confound. The size of the file is confounding the explanation that the FIX protocol is faster for sending financial data. In fact, it may depend on the size of the data file which protocol is faster. If large data sets had been used, the results may have indicated that SOAP was the faster protocol.

In a simple experimental group–control group experiment, if subjects in the experimental group are always administered treatment in the morning and subjects in the control group always receive the treatment in the afternoon, a systematic error occurs. In such a situation, time of day represents a

confound. In a training experiment the sources of constant error might be the persons who do the training (line or external specialists) or whether the training is conducted on the employees' own time or on company time.

EXTRANEOUS VARIABLES

Most business students realize that the marketing mix variables—price, product, promotion, and distribution—interact with uncontrollable forces in the market, such as economic variables, competitor activities, and consumer trends. Thus, many marketing experiments are subject to the effect of **extraneous variables**. Since extraneous variables can produce confounded results, they must be identified before the experiment if at all possible.

One issue with significant business and public policy implications is cigarette smoking. Does cigarette advertising cause young people to smoke? Although this is an often asked question, it is far from settled. One of the primary reasons for the inconclusiveness of this debate is the failure for most of the research to control for extraneous variables. For instance, consider a study in which two groups of U.S. high school students are studied over the course of a year. One is exposed to foreign television media in which American cigarettes are more often shown in a flattering and glamorous light. In fact, the programming includes cigarette commercials. The other group is a control group in which their exposure to media is not controlled. At the end of the year, the experimental group reports a greater frequency and incidence of cigarette smoking. Did the increased media exposure involving cigarettes cause smoking behavior?

While the result seems plausible at first, the careful researcher may ask the following questions:

- Was the demographic makeup of the two groups the same? While it is clear that the ages of the two groups are likely the same, it is well known that different ethnic groups have different smoking rates. Approximately 28 percent of all high school students report smoking, but the rate is higher among Hispanic teens, for example. Therefore, if one group contained more Hispanics, we might expect it to report different smoking rates than otherwise. Similarly, smoking varies with social class. Were the two groups comprised of individuals from comparable social classes?
- How did the control group fill the time consumed by the experimental group in being exposed to the experimental treatment? Could it be that it somehow dissuaded them from smoking? Perhaps they were exposed to media with more anti-smoking messages?
- Were the two groups of the same general achievement profiles? Those who are high in the need for achievement may be less prone to smoke than are other students.
- Although it is a difficult task to list all possible extraneous factors, some that even sound unusual can sometimes have an effect. For example, did the students have equally dispersed birthdays? Researchers have shown that smoking rates correspond to one's birthday, meaning that different astrological groups have different smoking rates.

Because an experimenter does not want extraneous variables to affect the results, he or she must eliminate or control such variables. It is always better to spend time thinking about how to control for possible extraneous variables before the experiment, since often there is nothing that can be done to salvage results if a confounding effect is identified after the experiment is conducted.

Demand Characteristics

What Are Demand Characteristics?

The term **demand characteristic** refers to an experimental design element that unintentionally provides subjects with hints about the research hypothesis. Researchers cannot reveal the research hypotheses to subjects before the experiment or else they can create a confounding effect. Think about the self-efficacy experiment. If the subjects learned that they were being intentionally given positive feedback to enhance their confidence and attitudes toward their job, the researcher would never be sure if their responses to the dependent variable were really due to the differences in the experimental stimuli or due to the fact that the subjects were trying to provide a "correct" response. Once subjects know the hypotheses, there is little hope that they will respond naturally. A confound may be created by knowledge of the experimental hypothesis. This particular type of confound is known as a **demand effect**. Demand characteristics make demand effects very likely.

Experimenter Bias and Demand Effects

Demand characteristics are aspects of an experiment that demand (encourage) that the subjects respond in a particular way. Hence, they are a source of systematic error. If participants recognize the experimenter's expectation or demand, they are likely to act in a manner consistent with the experimental treatment. Even slight nonverbal cues may influence their reactions.

Prominent demand characteristics are often presented by the person administering experimental procedures. If an experimenter's presence, actions, or comments influence the subjects' behavior or sway the subjects to slant their answers to cooperate with the experimenter, the experiment has introduced experimenter bias. When subjects slant their answers to cooperate with the experimenter, they are exhibiting behaviors that might not represent their behavior in the marketplace. For example, if subjects in an advertising experiment understand that the experimenter is interested in whether they changed their attitudes in accord with a given advertisement, they may answer in the desired direction. Acting in this manner reflects a demand effect rather than a true experimental treatment effect.

Hawthorne Effect

A famous management experiment illustrates a common demand characteristic. Researchers were attempting to study the effects on productivity of various working conditions, such as hours of work, rest periods, lighting, and methods of pay, at the Western Electric Hawthorne plant in Cicero, Illinois. The researchers found that workers' productivity increased whether the work hours were lengthened or shortened, whether lighting was very bright or very dim, and so on. The surprised investigators realized that the workers' morale was higher because they were aware of being part of a special experimental group. This totally unintended effect is now known as the **Hawthorne effect** because researchers realize that people will perform differently when they know they are experimental subjects.

If subjects in a laboratory experiment interact (i.e., are not relatively isolated), their conversations may produce joint decisions rather than a desired individual decision. For this reason, social interaction generally is restricted in laboratory experiments.

Reducing Demand Characteristics

Although it is practically impossible to eliminate demand characteristics from experiments, there are steps that can be taken to reduce them. Many of these steps make it difficult for subjects to know what the researcher is trying to find out. Some or all of these may be appropriate in a given experiment.

1. Use an experimental disguise.
2. Isolate experimental subjects.
3. Use a "blind" experimental administrator.
4. Administer only one experimental treatment level to each subject.

■ **EXPERIMENTAL DISGUISE**

Subjects participating in the experiment can be told that the purpose of the experiment is somewhat different than the actual purpose. Most often, they are simply told less than the complete "truth" about what is going to happen. In other cases, more deceit may be needed. For example, psychologists studying how much pain one person may be willing to inflict on another might use a ruse telling the subject that they are actually interested in the effect of pain on human performance. The researcher tells the subject to administer a series of questions to another person (who is actually a research assistant) and to provide them with an increasingly strong electric shock each time an incorrect answer is given. In reality, the real dependent variable has something to do with how long the actual subject will continue to administer shocks before stopping.

A **placebo** is an experimental deception involving a false treatment. A **placebo effect** refers to the corresponding effect in a dependent variable that is due to the psychological impact that goes along with knowledge of the treatment. A placebo is particularly important when the experimental variable involves physical consumption of some product. The placebo should not be different in any observable manner from the true treatment that is actually noticeable by the research subject. Assume a researcher is examining the ability of a special food additive to suppress appetite. The additive is a product that is supposed to be sprinkled on food before it is eaten. The experimental group would be given the actual product to test, while the control group would be given a placebo that looks exactly like the actual food additive but is actually an inert compound. Both groups are likely to show some difference in consumption compared to someone undergoing no effect. The difference in the actual experimental group and the placebo group would represent the true effect of the additive.

Placebo effects exist in marketing research. For example, when subjects are told that an energy drink is sold at a discount price, they believe it is significantly less effective than when it is sold at the regular, non-discounted price. Later, we will return to the ethical issues involved in experimental deception.

■ **ISOLATE EXPERIMENTAL SUBJECTS**

Researchers should minimize the extent to which subjects are able to talk about the experimental procedures with each other. Although it may be unintentional, discussion among subjects may lead them to guess the experimental hypotheses. For instance, it could be that different subjects received different treatments, which the subjects could discover if they talked to one another. The experimental integrity will be higher when each subject only knows enough to participate in the experiment.

■ **USE A "BLIND" EXPERIMENTAL ADMINISTRATOR**

When possible, the people actually administering the experiment may not be told the experimental hypotheses. The advantage is that if they do not know what exactly is being studied, then they are less likely to give off clues that result in demand effects. Like the subjects, when there is some reason to expect that their knowledge may constitute a demand characteristic, administrators best know only enough to do their job.

■ **ADMINISTER ONLY ONE EXPERIMENTAL CONDITION PER SUBJECT**

When subjects know more than one experimental treatment condition, they are much more likely to guess the experimental hypothesis. So, even though there are cost advantages to administering multiple treatment levels to the same subject, it should be avoided when possible. For example, in the self-efficacy experiment, if the subjects were asked to complete a questionnaire regarding their self-

confidence in doing their job and their job satisfaction, then again asked to respond to the same questions after the personal interview, and then again after each of the three letters giving them positive feedback, they are very likely to guess that they are intentionally being given the feedback to enhance their self-efficacy.

Establishing Control

The major difference between experimental research and descriptive research is an experimenter's ability to control variables by either holding conditions constant or manipulating the experimental variable. If the color of beer causes preference, a brewery experimenting with a new clear beer must determine the possible extraneous variables other than color that may affect an experiment's results and attempt to eliminate or control those variables. Marketing theory tell us that brand image and packaging design are important factors in beer drinkers' reactions. Therefore, the researcher may wish to control the influence of these variables. He or she may eliminate these two extraneous variables by packaging the test beers in plain brown packages without any brand identification.

When extraneous variables cannot be eliminated, experimenters may strive for **constancy of conditions**. This means that subjects in all experimental groups are exposed to identical conditions except for the differing experimental treatments. Random assignment and the principle of matching discussed earlier help make sure that constancy is achieved.

A supermarket experiment involving shelf space shows the care that must be taken to hold all factors constant. The experiment required that all factors other than shelf space be kept constant throughout the testing period. In all stores, the shelf level that had existed before the test began was to be maintained throughout the test period. Only the *amount* of shelf space (the treatment) was changed. One problem involved store personnel accidentally changing shelf level when stocking the test products. This deviation from the constancy of conditions was minimized by auditing each store four times a week. In this way, any change could be detected in a minimum amount of time. The experimenter personally stocked as many of the products as possible, and the cooperation of stock clerks also helped reduce treatment deviations.

If an experimental method requires that the same subjects be exposed to two or more experimental treatments, an error may occur due to the *order of presentation.* For instance, if subjects are examining the effects of different levels of graphical interface on video game enjoyment, and they are asked to view each of four different levels, the order in which they are presented may influence enjoyment. Subjects might prefer one level simply because it follows a very poor level. **Counterbalancing** attempts to eliminate the confounding effects of order of presentation by requiring that one-fourth of the subjects be exposed to treatment A first, one-fourth to treatment B first, one-fourth to treatment C first, and finally one-fourth to treatment D first. Likewise, the other levels are counterbalanced so that the order of presentation is rotated among subjects. It is easy to see where counterbalancing is particularly important for experiments such as taste tests, where the order of presentation may have significant effects on consumer preference.

Problems Controlling Extraneous Variables

In many experiments it is not always possible to control every potential extraneous variable. For example, competitors may bring out a product during the course of a test-market. This form of competitive interference occurred in a Boston test-market for Anheuser-Busch's import beer, Wurzburger Hofbrau. During the test, Miller Brewing Company introduced its own brand, Munich Oktoberfest, and sent eight salespeople out to blitz the Boston market. A competitor who learns of a test-market experiment may knowingly change its prices or increase advertising to confound the test results. This brings us to ethical issues in experimentation.

Ethical Issues in Experimentation

Ethical issues with business research were discussed in Chapter 5. There, the question of deception was raised. Although deception is necessary in most experiments, when subjects can be returned to their prior condition through debriefing, then the experiment is probably consistent with high moral standards. If subjects might be injured significantly or truly psychologically harmed, debriefing will not return them to their formal condition and the experiment should not be undertaken. Therefore, some additional commentary on debriefing is warranted. Debriefing experimental subjects by communicating the purpose of the experiment and the researcher's hypotheses is expected to counteract negative effects of deception, relieve stress, and provide an educational experience for the subject.

Additionally, there is the issue of test-markets and efforts extended toward interfering with a competitor's test-market. When a company puts a product out for public consumption, they should be aware that competitors may also now freely consume the product. When attempts to interfere with a test-market are aimed solely at invalidating test results or they are aimed at infringing on some copyright protection, those acts are ethically questionable.

Practical Experimental Design Issues

Basic versus Factorial Experimental Designs

In *basic experimental designs* a single independent variable is manipulated to observe its effect on a single dependent variable. Our example of the computer communication protocols falls into this category—one independent variable (the two protocols) was examined and one dependent measure (speed of the data transfer) was assessed. However, we know that most business situations are much more complex and multiple independent and dependent variables are possible. Our self efficacy experiment illustrated this as both the treatment and the length of time as an employee were independent variables and multiple dependent variables were examined. In a complex marketing experiment, multiple dependent variables such as sales, product usage, and preference are influenced by several factors. The simultaneous change in independent variables such as price and advertising may have a greater influence on sales than if either variable is changed alone. In job satisfaction studies, we know that no one thing totally determines job satisfaction. Salary, opportunities for advancement, the pleasantness of the workplace, interactions with colleagues, and many more factors all combine and interact to determine how satisfied employees are with their job. *Factorial experimental designs* are more sophisticated than basic experimental designs and allow for an investigation of the interaction of two or more independent variables.

Laboratory Experiments

A business experiment can be conducted in a natural setting (a field experiment) or in an artificial setting (a laboratory experiment). In social sciences, the actual laboratory may be a behavioral lab, which is somewhat like a focus group facility. However, it may simply be a room or classroom dedicated to collecting data, or it can even take place in one's home. In a **laboratory experiment** the researcher has more complete control over the research setting and extraneous variables. Our example of the financial protocol experiment illustrates the benefits of a laboratory setting. The researchers were able to control for many factors, such as the size of the data file, the models of the computers, the Internet line, and so forth. This enhanced their confidence in establishing that the differences noted in speed were due to the different protocols. However, the researchers were not able to determine how the protocols compared when used in the field, on various computers, with a variety of file sizes, and under differing "real-world" circumstances.

Field Experiments

Field experiments are research projects involving experimental manipulations that are

implemented in a natural environment. They can be useful in fine-tuning managerial strategies and tactical decisions. Our self-efficacy study is an example of a field experiment. Rather than bring subjects into an artificial setting and trying to manipulate their self-efficacy and then measure their perceptions of job satisfaction, the researchers took their experiment to the field and used actual employees, which were provided feedback from their supervisors. In the marketing discipline, test-markets are good examples of field experiments. Betty Crocker's Squeezit (a 10 percent fruit juice drinks in a squeeze bottle) was so successful in a test-market that production could not keep up with demand. As a result, the product's national introduction was postponed until production capacity could be increased. McDonald's conducted a field experiment testing the Triple Ripple, a three-flavor ice cream product. The product was dropped because the experiment revealed distribution problems reduced product quality and limited customer acceptance. In the distribution system the product would freeze, defrost, and refreeze. Solving the problem would have required each McDonald's city to have a local ice cream plant with special equipment to roll the three flavors into one. While a laboratory experiment might have shown tremendous interest, a naturalistic setting for the experiment helped McDonald's executives realize the product was impractical.

Within-Subjects and Between-Subjects Designs

A basic question faced by the researchers involves how many treatments a subject should receive. For economical reasons, the researcher may wish to apply multiple treatments to the same subject. Thus, multiple observations on the dependent variable can be obtained from a single subject. Such a design is called a **within-subjects design**. Within-subjects designs involve repeated measures because with each treatment the same subject is measured.

In contrast, the researcher could decide that each person will receive only one treatment combination. This is referred to as a **between-subjects design**. Each dependent variable is measured only once for every subject. Between-subjects designs are usually advantageous even though they are typically more costly. The validity of between-subjects designs is usually higher since applying only one treatment combination to each subject greatly reduces the possibility of demand characteristics. In addition, as we will see later, statistical analysis of between-subjects designs are simpler than within-subjects designs. This also means the results are easier to report and explain to management.

Issues of Experimental Validity

An experiment's quality is judged by two types of validity. These are known as internal and external validity.

1-Internal Validity

Internal validity exists to the extent that an experimental variable is truly responsible for any variance in the dependent variable. In other words, does the experimental manipulation truly cause changes in the specific outcome of interest? If the observed results were influenced or confounded by extraneous factors, the researcher will have problems making valid conclusions about the relationship between the experimental treatment and the dependent variable.

Thus, a lab experiment enhances internal validity because it maximizes control of outside forces. If we wish to know whether certain music causes increased productivity among workers, we may set up a task in a room with different music piped in (our experimental manipulation), but with the temperature, lighting, density, other sounds, and any other factors all controlled, which would be difficult or impossible to control outside of a lab environment. If the only thing that varies from subject to subject is the music, then we can safely say that any differences in performance must be attributable to human reactions to the music. Our opening example of the protocol experiment focused on maximizing internal validity. By testing the protocols in a lab setting, the researchers were able to control extraneous variables such as differences in computing hardware, network issues, and so forth.

a-MANIPULATION CHECKS

Internal validity depends in large part on successful manipulations. Manipulations should be carried out in a way that the independent variable differs over meaningful levels. If the levels are too close together, the experiment may lack the power necessary to observe differences in the dependent variable. In a pricing experiment, it may be that manipulating the price of an automobile over two levels, $24,600 and $24,800, would not be successful in creating truly different price categories. Respondents might not perceive the differences or experience any reaction to such a slight deviation. The validity of manipulations can often be determined with a **manipulation check**. If a drug is administered in different dosages that should affect blood sugar levels, the researcher could actually measure blood sugar level after administering the drug to make sure that the dosages were different enough to produce a change in blood sugar.

In business research, the manipulation check is often conducted by asking a survey question or two. In the pricing example above, subjects may be asked a question about how low they believe the price of the car to be. A valid manipulation would produce substantially different average responses to that question in a "high" and "low" price group. In our self-efficacy example, the researchers were interested in the impact increased self-efficacy (the independent variables) had on job satisfaction (the dependent variable). The experimental manipulation was the positive feedback the subjects were given. The manipulation check was a series of questions assessing the subject's self-efficacy. Did the positive feedback actually increase self-efficacy? If it did, then the researchers could examine the other relationships of interest. However, if self-efficacy did not increase, then the researchers would have to reconsider their manipulation and find another way to enhance self-efficacy to carry out the study. Manipulation checks should always be administered after dependent variables in self-response format experiments. This keeps the manipulation check item from becoming a troublesome demand characteristic. Extraneous variables can jeopardize internal validity. The six major ones are *history, maturation, testing, instrumentation, selection,* and *mortality.*

b-HISTORY

A **history effect** occurs when some change other than the experimental treatment occurs during the course of an experiment that affects the dependent variable. A common history effect occurs when competitors change their marketing strategies during a test marketing experiment. Another example would be if some of our subjects in the self-efficacy exam are offered a position by another firm. A different job offer may affect several of the dependent measures in the study.

History effects are particularly prevalent in repeated measures experiments that take place over an extended time. If we wanted to assess how much a change in recipe improves individual subjects' consumption of a food product, we would first measure their consumption and then compare it with consumption after the change. Since several weeks may pass between the first and second measurement, there are many things that could occur that would also influence subjects' diets. Although it would be extreme, examining the effect of a dietary supplement on various health-related outcomes may require that a subject be confined during the experiment's course. This may take several weeks. Without confining the subject in something like a hospital setting, there would be little way of controlling food and drink consumption, exercise activities, and other factors that may also affect the dependent variables.

A special case of the history effect is the **cohort effect**, which refers to a change in the dependent variable that occurs because members of one experimental group experienced different historical situations than members of other experimental groups. For example, groups of managers used as subjects may be in different cohorts because one group encountered different experiences over the course of an experiment.

c-MATURATION

Maturation effects are effects that are a function of time and the naturally occurring events that coincide with growth and experience. Experiments taking place over longer time spans may see

lower internal validity as subjects simply grow older or more experienced. For example, our self-efficacy study shows that job satisfaction seems to decline with time. Conversely, job skill tends to increase over time. Suppose an experiment were designed to test the impact of a new compensation program on sales productivity. If this program were tested over a year's time, some of the salespeople probably would mature as a result of more selling experience and gain increased knowledge and skill. Their sales productivity might improve because of their knowledge and experience rather than the compensation program.

d-TESTING

Testing effects are also called *pretesting effects* because the initial measurement or test alerts or primes subjects in a way that affects their response to the experimental treatments. Testing effects only occur in a before-and-after study. A before-and-after study is one requiring an initial baseline measure be taken before an experimental treatment is administered. So, before-and- after experiments are a special case of a repeated measures design. For example, students taking standardized achievement and intelligence tests for the second time usually do better than those taking the tests for the first time. The effect of testing may increase awareness of socially appropriate answers, increase attention to experimental conditions (that is, the subject may watch more closely), or make the subject more conscious than usual of the dimensions of a problem.

e-INSTRUMENTATION

A change in the wording of questions, a change in interviewers, or a change in other procedures used to measure the dependent variable causes an **instrumentation effect**, which may jeopardize internal validity. Sometimes instrumentation effects are difficult to control. For example, if the same interviewers are used to ask questions for both before and after measurement, some problems may arise. With practice, interviewers may acquire increased skill in interviewing, or they may become bored and decide to reword the questionnaire in their own terms. To avoid this problem, new interviewers could be hired. But this introduces another set of issues as different individuals are also a source of extraneous variation. There are numerous other sources of instrument decay or variation. Again, instrumentation effects are problematic with any type of repeated measures design.

f-SELECTION

The **selection effect** is a sample bias that results from differential selection of respondents for the comparison groups, or sample selection error, discussed earlier. Researchers must make sure the characteristics of the research subjects accurately reflect the population of relevance. Furthermore, the key characteristics of the subjects must be distributed in such a way to create equal groups. That is, the subjects in the experimental and control groups, or in different experimental cells, must be equal across all variables of interest and those that could affect the dependent measure.

g-MORTALITY

If an experiment is conducted over a period of a few weeks or more, some sample bias may occur due to the **mortality effect (sample attrition)**. Sample attrition occurs when some subjects withdraw from the experiment before it is completed. Mortality effects may occur if subjects drop from one experimental treatment group disproportionately than from other groups. Consider a sales training experiment investigating the effects of close supervision of salespeople (high pressure) versus low supervision (low pressure). The high-pressure condition may misleadingly appear superior if those subjects who completed the experiment did very well. If, however, the high-pressure condition caused more subjects to drop out than the other conditions, this apparent superiority may be due to the fact that only very determined and/or talented salespeople stuck with the program. Similarly, in the self-efficacy study, accountants that did not feel commitment to the organization and maintain a high level of job satisfaction may have left the organization before the final measures.

2-External Validity

External validity is the accuracy with which experimental results can be generalized beyond the experimental subjects. External validity is increased when the subjects comprising the sample truly represent the population of interest and when the results extend to other market segments or groups of people. The higher the external validity, the more researchers and managers can count on the fact that any results observed in an experiment will also be seen in the "real world" (financial market, workplace, sales floor, and so on).

For instance, to what extent would results from our protocol experiment, which represents a simulated financial market data exchange, transfer to a real-world trading situation? Would the FIX protocol prove to be faster across computer systems, Internet line transfer speeds, and different traders? Would increases in self- efficacy enhance the job satisfaction of retail store workers, salespeople, or human resource managers as it did for accountants? Can one extrapolate the results from a tachistoscope to an in- store shopping situation? Lab experiments, such as the protocol experiment, are associated with low external validity because the limited set of experimental conditions, holding all else constant, do not adequately represent all the influences existing in the real world. In other words, the experimental situation may be too artificial. When a study lacks external validity, the researcher will have difficulty repeating the experiment with any change in subjects, settings, or time.

STUDENT SUBJECTS

Basic researchers often use college students as experimental subjects. Convenience, time, money, and a host of other practical considerations often result in students being used as research subjects. This practice is widespread in academic studies. Some evidence shows that students are quite similar to household consumers, but other evidence indicates that they do not provide sufficient external validity to represent most consumer or employee groups. This is particularly true when students are used as substitutes or surrogates for businesspeople. The issue of external validity should be seriously considered because the student population is likely to be atypical. Students are easily accessible, but they often are not representative of the total population. This is not always the case, however, and when behaviors are studied for which students have some particular expertise (the purchase of relevant products such as MP3 players or job search skills), then they are certainly appropriate. For instance, the Research Snapshot "Does Promotion Cause Intoxication?" on page 262 is an example where students are very appropriate research subjects.

Trade-Offs Between Internal and External Validity

Naturalistic field experiments tend to have greater external validity than artificial laboratory experiments. Researchers often must trade internal validity for external validity. A researcher who wishes to test advertising effectiveness by manipulating treatments via a split-cable experiment has the assurance that the advertisement will be viewed in an externally valid situation, the subjects' homes. However, the researcher has no assurance that some interruption (for example, the telephone ringing, a child calling, or a pot boiling over on the stove) will not have some influence that will reduce the internal validity of the experiment. Laboratory experiments with many controlled factors usually are high in internal validity, while field experiments generally have less internal validity but greater external validity. Typically, it is best to establish internal validity first, and then focus on external validity. Thus, results from lab experiments would be followed up with some type of field test.

Classification of Experimental Designs

An experimental design may be compared to an architect's plans for a building. The basic requirements for the structure are given to the architect by the prospective owner. Several different

plans may be drawn up as options for meeting the basic requirements. Some may be more costly than others. One may offer potential advantages that another does not. There are various types of experimental designs. If only one variable is manipulated, the experiment has a **basic experimental design**. If the experimenter wishes to investigate several levels of the independent variable (for example, four different employee salary levels) or to investigate the interaction effects of two or more independent variables (salary level and retirement package), the experiment requires a *complex,* or *statistical,* experimental design.

Symbolism for Diagramming Experimental Designs

The work of Campbell and Stanley has helped many students master the subject of basic experimental designs. The following symbols will be used in describing the various experimental designs:

X = exposure of a group to an experimental treatment
*O = observation or measurement of the dependent variable; if more than one observation or measurement is taken, subscripts (that is, O*1*, O*2*, etc.) indicate temporal order*
R = random assignment of test units; R symbolizes that individuals selected as subjects for the experiment are randomly assigned to the experimental groups

Three Examples of Quasi-Experimental Designs

Quasi-experimental designs do not involve random allocation of subjects to treatment combinations. In this sense, they do not qualify as true experimental designs because they do not adequately control for the problems associated with loss of internal validity. However, sometimes quasi-experimental designs are the only way to implement a study.

1-ONE SHOT DESIGN

The one-shot design, or *after-only design,* is diagrammed as follows:

$$X\ O_1$$

Suppose that during a very cold winter an automobile dealer finds herself with a large inventory of cars. She decides to experiment for the month of January with a promotional scheme. She offers a free trip to New Orleans with every car sold. She experiments with the promotion (X = experimental treatment) and measures sales (O_1= measurement of sales after the treatment is administered). This one-shot design is a case study of a research project fraught with problems. Subjects or test units participate because of voluntary self-selection or arbitrary assignment, not because of random assignment. The study lacks any kind of comparison or any means of controlling extraneous influences. There should be a measure of what will happen when the test units have not been exposed to X to compare with the measures of when subjects have been exposed to X. The one-shot experimental design commonly suffers from most of the threats to internal validity discussed above. Nevertheless, under certain circumstances, it is the only viable choice.

2-ONE GROUP PRETEST POSTTEST DESIGN

Suppose a real estate franchiser wishes to provide a training program for franchisees. If the franchiser measures subjects' knowledge of real estate selling before (O1) they are exposed to the experimental treatment (X) and then measures real estate selling knowledge after (O2) they are exposed to the treatment, the design will be as follows:

$$O_1\ X\ O_2$$

In this example the trainer is likely to conclude that the difference between O_2 and O_1 ($O_2 - O_1$) is the measure of the influence of the experimental treatment. This one-group pretest–posttest design offers a comparison of the same individuals before and after training. Although this is an improvement over the one-shot design, this research still has several weaknesses that may jeopardize

internal validity. For example, if the time lapse between O_1 and O_2 was a period of several months, the trainees may have matured as a result of experience on the job (maturation effect). History effects—such as a change in interest rates—may also influence the dependent measure in this design. Perhaps some subjects dropped out of the training program (mortality effect). The effect of testing may also have confounded the experiment. Although this design has a number of weaknesses, it is commonly used in business research. Remember, the cost of the research is a consideration in most business situations. While there will be some problems of internal validity, the researcher must always take into account questions of time and cost.

3-STATIC GROUP DESIGN

In a static group design, each subject is identified as a member of either an experimental group or a control group (for example, exposed or not exposed to a training program). The experimental group is measured after being exposed to an experimental treatment and the control group is measured without having been exposed to this experimental treatment:

Experimental group:	*X*	O_1
Control group:		O_2

The results of the static group design are computed by subtracting the observed results in the control group from those in the experimental group ($O_1 - O_2$). A major weakness of this design is its lack of assurance that the groups were equal on variables of interest before the experimental group received the treatment. If entry into either group was voluntary, systematic differences between the groups could invalidate the conclusions about the effect of the treatment. For example, if the real estate franchisor mentioned above asked her franchisees who would like to attend the training program, we have no way of knowing if those who chose to attend are the same as those who did not. Random assignment of subjects may eliminate problems with group differences. If groups are established by the experimenter rather than existing as a function of some other causation, the static group design is referred to as an *after-only design with control group.* On many occasions, an after-only design is the only possible option. This is particularly true when conducting use tests for new products or brands. Cautious interpretation and recognition of the design's shortcomings may enhance the value of this design.

Three Alternative Experimental Designs

In a formal scientific sense, the three designs just discussed are not pure experimental designs. Subjects for the experiments were not selected from a common pool of subjects and randomly assigned to one group or another. In the following discussion of three basic experimental designs, the symbol to the left of the diagram indicates that the first step in a true experimental design is the random assignment of subjects.

1-PRETEST POSTTEST CONTROL GROUP DESIGN (BEFORE AFTER WITH CONTROL) A pretest–posttest control group design, or *before–after with control group design,* is the classic experimental design:

Experimental group:	R	O_1	*X*	O_2
Control group:	R	O_3		O_4

As the diagram indicates, the subjects in the experimental group are tested before and after being exposed to the treatment. The control group is tested at the same two times as the experimental group, but subjects are not exposed to the experimental treatment. This design has the advantages of the before–after design with the additional advantages gained by its having a control group. The effect of the experimental treatment equals:

$$(O2 - O1) - (O4 - O3)$$

It is important to note that we expect $O1 = O3$. One of the threats we discussed to internal validity was selection and the assumption of equal groups. If the two groups are not equal at the beginning of the experiment, the study has a fatal flaw and the researchers should start over. Let's assume there is brand awareness among 20 percent of the subjects ($O1$ = 20 percent, $O3$ = 20 percent) before an advertising treatment and then 35 percent awareness in the experimental group ($O2$ = 35 percent) and 22 percent awareness in the control group ($O4$ = 22 percent) after exposure to the treatment, the treatment effect equals 13 percent:

$$(0.35 - 0.20) - (0.22 - 0.20) = (0.15) - (0.02) = 0.13 \textit{ or } 13\%$$

Not only are the groups assumed to be equal at the beginning, but the effect of all extraneous variables is assumed to be the same on both the experimental and the control groups. For instance, since both groups receive the pretest, no difference between them is expected for the pretest effect. This assumption is also made for effects of other events between the before and after measurements (history), changes within the subjects that occur with the passage of time (maturation), testing effects, and instrumentation effects. In reality there may be some differences in the sources of extraneous variation. Nevertheless, in most cases assuming that the effect is approximately equal for both groups is reasonable. However, a testing effect is possible when subjects are sensitized to the subject of the research. This is analogous to what occurs when people learn a new vocabulary word. Soon they discover that they notice it much more frequently in their reading. In an experiment the combination of being interviewed on a subject and receiving the experimental treatment might be a potential source of error. For example, a subject exposed to a certain advertising message in a split-cable experiment might say, "Ah, there is an ad about the product I was interviewed about yesterday!"

2-POSTTEST ONLY CONTROL GROUP DESIGN (AFTER ONLY WITH CONTROL)

In some situations pretest measurements are impossible. In other situations selection error is not anticipated to be a problem because the groups are known to be equal. The posttest-only controls group design, or *after-only with control group design,* is diagrammed as follows:

Experimental group:	R	*X*	*O*1
Control group:	R		*O*2

The effect of the experimental treatment is equal to $O1 - O2$.

Suppose the manufacturer of an athlete's-foot remedy wishes to demonstrate by experimentation that its product is better than a competing brand. No pretest measure about the effectiveness of the remedy is possible. The design is to randomly select subjects, perhaps students, who have contracted athlete's foot and randomly assign them to the experimental or the control group. With only the posttest measurement, the effects of testing and instrument variation are eliminated. Furthermore, researchers make the same assumptions about extraneous variables described above—that is, that they operate equally on both groups, as in the before–after with control group design.

3-COMPROMISE DESIGNS

True experimentation is often simply not possible. The researcher may compromise by approximating an experimental design. A compromise design is one that falls short of assigning subjects or treatments randomly to experimental groups. Consider a situation in which a researcher would ideally implement a pretest–posttest control group design to study the effect of training on employee performance. In this case, subjects may not be able to be assigned randomly to the experimental and control group because the researcher cannot take workers away from their work groups. Thus, one entire work group is used as the experimental group and a separate work group is used as a control group. The researcher has no assurance that the groups are equivalent. The situation has forced a compromise to experimental integrity. The alternative to the compromise design when

random assignment of subjects is not possible is to conduct the experiment *without* a control group. Generally this is considered a greater weakness than using groups that have already been established. When the experiment involves a longitudinal study, circumstances usually dictate a compromise with true experimentation.

Time Series Designs

Many experiments may be conducted in a short period of time (a few hours, a week, or a month). However, a business experiment investigating long-term strategic and/or structural changes may require a **time series design**. Time series designs are quasi-experimental because they generally do not allow the researcher full control over the treatment exposure or influence of extraneous variables. When experiments are conducted over long periods of time, they are most vulnerable to history effects due to changes in population, attitudes, economic patterns, and the like. Although seasonal patterns and other exogenous influences may be noted, the experimenter can do little about them when time is a major factor in the design. Political tracking polls provide an example. A pollster normally uses a series of surveys to track candidates' popularity. Consider the candidate who plans a major speech (the experimental treatment) to refocus the political campaign. The simple time series design can be diagrammed as follows:

$$O1 \quad O2 \quad O3 \quad X \quad O4 \quad O5 \quad O6$$

Several observations have been taken to identify trends before the speech (X) is given. After the treatment has been administered, several observations are made to determine if the patterns *after* the treatment are similar to those *before.* If the longitudinal pattern shifts after the political speech, the researcher may conclude that the treatment had a positive impact on the pattern. Of course, this time series design cannot give the researcher completes assurance that the treatment caused the change in the trend, rather than some external event. Problems of internal validity are greater than in more tightly controlled before-and-after designs for experiments of shorter duration. One unique advantage of the time series design is its ability to distinguish temporary from permanent changes. Exhibit below shows some possible outcomes in a time series experiment.

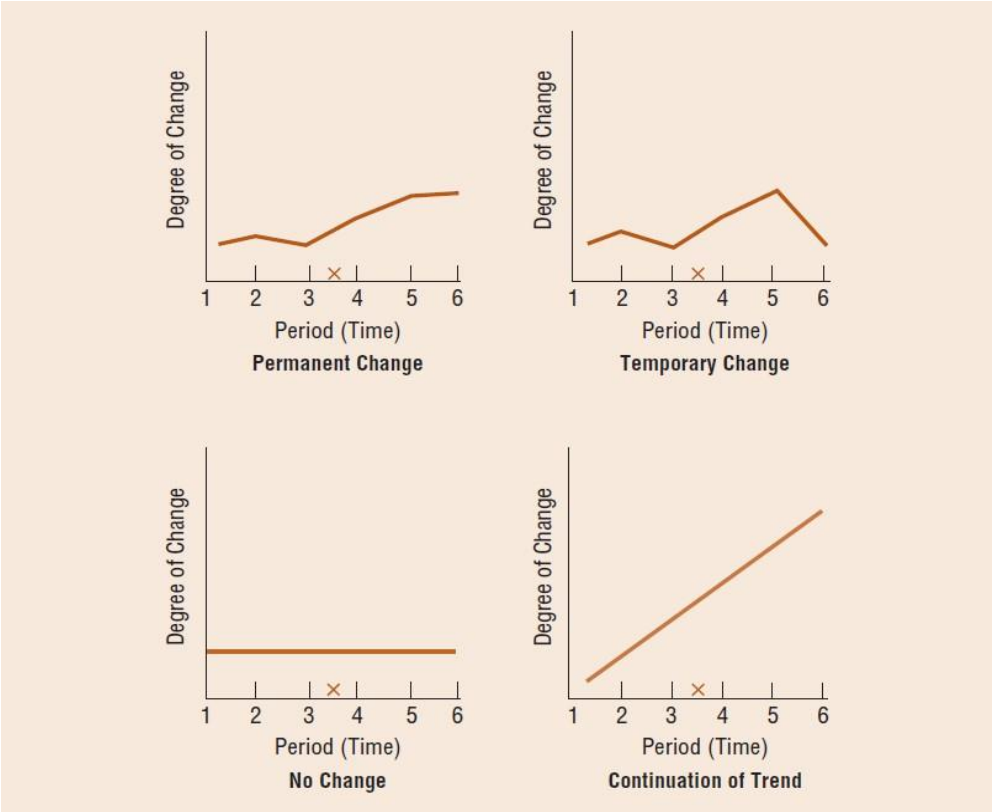


EXHIBIT Selected Time Series Outcomes

Complex Experimental Designs

The previous discussion focused on simple experimental designs—experiments manipulating a single variable. Here, the focus shifts to more complex experimental designs involving multiple experimental variables. Complex experimental designs are statistical designs that isolate the effects of confounding extraneous variables or allow for manipulation of more than one independent variable in the experiment. *Completely randomized designs, randomized block designs, and factorial designs* are covered in the following section.

COMPLETELY RANDOMIZED DESIGN

A **completely randomized design** is an experimental design that uses a random process to assign subjects to treatment levels of an experimental variable. Randomization of experimental units is the researcher's attempt to control extraneous variables while manipulating potential causes. A one-variable experimental design can be completely randomized, so long as subjects are assigned in a random way to a particular experimental treatment level. Consider a financial institution that wants to increase their response to credit card offers. An experiment is constructed to examine the effects of various incentives on the percentage of potential customers that apply for a credit card with the institution. Thus, the experimental variable is the incentive. This can be manipulated over three treatment levels:

1. No incentive to the control group
2. No interest for the first 90 days with an approved application
3. A free MP3 player with an approved application

The financial institution rents a mailing list of 15,000 prospects. This sample frame is divided into three groups of 5,000 each ($n1 + n2 + n3 = 15{,}000$). A random number process could be used to assign subjects to one of the three groups. Suppose each of the 15,000 subjects is assigned a number ranging from 1 to 15,000. If a random number is selected between 1 and 15,000 (i.e., 1,201), that person can be assigned to the first group, with every third person afterward and before also assigned to the first group (1,204, 1,207, 1,210 . . . all the way back to 1,198). The process can be repeated with the remaining 10,000 subjects by selecting a random number between 1 and 10,000 and then selecting every other subject. At this point, only 5,000 subjects remain and will comprise the third group. All 15,000 subjects are now assigned to one of three groups. Each group corresponds to one of the three levels of incentive. A variable representing which group a subject belongs to becomes the independent variable. The dependent variable is measured for each of the three treatment groups and the number of respondents to the offer is determined. The analysis would compare differences across the number of respondents for each of the three treatment levels.

RANDOMIZED BŁOCK DESIGN

The **randomized-block design** is an extension of the completely randomized design. A form of randomization is used to control for *most* extraneous variation; however, the researcher has identified a single extraneous variable that might affect subjects' responses systematically. The researcher will attempt to isolate the effects of this single variable by blocking out its effects.

A blocking variable is a categorical variable that is expected to be associated with different values of a dependent variable for each group. Sex is a common blocking variable. Many potential dependent variables are expected to be different for men and women. For instance, work-family conflict—conflict between the obligations a person has to their family and with their work commitments—has been found to differ between women and men. So, if a researcher is studying how salary and length of vacation time affects employee job satisfaction, they may want to also record a person's sex and include it as an extra explanatory variable over and above the experimental variable's salary and vacation time. The concept of a blocking variable was introduced in the self-efficacy study where the researchers "blocked" on length of time the subjects had been an employee (new versus current).

The term *randomized block* originated in agricultural research that applied several levels of a treatment variable to each of several blocks of land. Systematic differences in agricultural yields due to the quality of the blocks of land may be controlled in the randomized-block design. In business research, the researcher may wish to isolate block effects such as bank branch territories, job work units, or employee tenure, and so on. Suppose that a manufacturer of Mexican food is considering two packaging alternatives. Marketers suspect that geographic region might confound the experiment. They have identified three regions where attitudes toward Mexican food may differ (the Southwest, the Midwest, and the Atlantic Coast).

In a randomized-block design, each block must receive every treatment level. Assigning treatments to each block is a random process. In this example the two treatments will be randomly assigned to different cities within each region. Sales results such as those in Exhibit below might be observed. The logic behind the randomized block design is similar to that underlying the selection of a stratified sample rather than a simple random one. By isolating the block effects, one type of extraneous variation is partitioned out and a more efficient experimental design therefore results. This is because experimental error is reduced with a given sample size.

	Percentage Who Purchase Product			
Treatment	**Southwest**	**Midwest**	**Atlantic Coast**	**Mean for Treatments**
Package A	14.0% (Phoenix)	12.0% (St. Louis)	7.0% (Boston)	11.0%
Package B	16.0% (Albuquerque)	15.0% (Peoria)	10.0% (New York)	13.6%
Mean for cities	15.0%	13.5%	8.5%	

EXHIBIT **Randomized Block Design**

FACTORIAL DESIGNS

Suppose a human resource manager believes that an experiment that manipulates the level of salary offered is useful, but too limited. The recruiters for the firm have been visiting college campuses and know that graduates seeking jobs are concerned about salary, but they are also concerned about the number of vacation days they will receive. However, the level of salary and actual number of vacation days needs to determined. Thus, an experiment to assess this requires more than one independent variable be incorporated into the research design. Even though the single-factor experiments considered so far may have one specific variable blocked and other confounding sources controlled, they are still limited. A **factorial design** allows for the testing of the effects of two or more treatments (factors) at various levels.

Main effects are differences (in the dependent variable) between treatment levels. Interactions produce differences (in the dependent variable) between experimental cells based on combinations of variables. In the self-efficacy example, we learned that the experimental treatment had a stronger effect on the current employees than the new employees.

	Vacation Days	
Salary	**10 Days**	**14 Days**
$37,500	Cell 1	Cell 4
$40,000	Cell 2	Cell 5
$42,500	Cell 3	Cell 6

To further explain the terminology of experimental designs, let us develop the recruiting experiment more fully. The human resource manager wants to measure the effect of the salary and vacation days on the percentage of job offers accepted. Exhibit 12.10 indicates three treatment levels of salary offered ($37,500, $40,000, and $42,500) and two levels of vacation time (10 days and 14 days). The table shows that every combination of treatment level requires a separate experimental group. In this experiment, with three levels of salary and two levels of vacation, we have a 3 × 2 (read “three by two”)

factorial design because the first factor (the salary variable) is varied in three ways and the second factor (the location variable) is varied in two ways. A 3 × 2 design requires six cells, or six experimental groups (3 × 2 = 6). If the subjects each receive only one combination of experimental variables, then we use the term 3 × 2 between-subjects design to describe the experiment.

The number of treatments (factors) and the number of levels of each treatment identify the factorial design. A 3 × 3 design means there are two factors, each having three levels; a 2 × 2 × 2 design has three factors, each having two levels. The treatments need not have the same number of levels; for example, a 3 × 2 × 4 factorial design is possible. The important idea is that in a factorial experiment, each treatment level is combined with every other treatment level.

In addition to the advantage of investigating two or more independent variables simultaneously, factorial designs allow researchers to measure interaction effects. In a 2 × 2 experiment the interaction is the effect produced by treatments A and B combined. If the effect of one treatment differs at various levels of the other treatment, interaction occurs.

To illustrate the value of a factorial design, suppose a researcher is comparing two magazine ads. The researcher is investigating the believability of ads on a scale from 0 to 100 and wishes to consider the sex of the reader as a blocking factor. The experiment has two independent variables: sex and ads. This 2 × 2 factorial experiment permits the experimenter to test three hypotheses.

Two hypotheses examine the main effects:

- Advertisement A is more believable than ad B.
- Men believe advertisements more than women.
- Advertisement A is more believable than ad B among women, but ad B is more believable than ad A among men.

A high score indicates a more believable ad. Exhibit 12.11 on the next page shows that the mean believability score for both sex is 65. This suggests that there is no main sex effect. Men and women evaluate believability of the advertisements equally. The main effect for ads indicates that ad A is more believable than ad B (70 versus 60), supporting the first hypothesis. However, if we inspect the data and look within the levels of the factors, we find that men find ad B more believable and women find ad A more believable.