

# >snapshot

Launching a new television series can be gut-wrenching. Networks spend millions developing new shows only to have them rejected in one or two episodes by the viewers. But some shows have an edge—or more specifically a dial. MSInteractive invented its Perception Analyzer™ technology more than 20 years ago. Each spring in cities across the country, samples of viewers watch pilot episodes of potential new shows, responding with Perception Analyzer's handheld dial to everything they see. Don't like a character? The viewer turns the dial to a number below 50 (the neutral point). A particular bit makes the viewer laugh, and he or she turns the dial to 80. "We collect data every second on what the viewer is feeling," shared David Paull, director of business development and marketing. "At the end of a 30-minute sitcom pilot, we have 1,800 data points for every member of the sample." To familiarize the sample with the dial, each participant is asked a series of questions up front. Some researchers will quiz participants on discrete choices like their gender (dial 1 for male, 2 for female); others will

query them on an intensity scale using 0 to 100 for a question like "How much of a fan are you of situation comedies?" All the data can be summarized on bar charts and tables, but the power of Perception Analyzer™ is its ability to superimpose a continuously generated line graph over the show itself. "The network can see precisely what appealed—or didn't appeal—not only to the audience as a whole, but also to specific demographic segments." Moderators in follow-up focus groups or personal interviews use the data to formulate questions to probe for understanding of participant reactions. "This approach generates real-time, anonymous, and accurate data to help researchers ask the right questions to gain true market insight." In the tested pilot episode of *ER*, actress Julianna Margulies commits suicide and dies. However, by the time this award-winning show actually aired, Julianna is rescued by George Clooney's character and proceeds to enjoy five seasons of health, including the birth of twins. "Understanding how the audience felt about Julianna's character was the insight that led to the change." [www.perceptionanalyzer.com](http://www.perceptionanalyzer.com)

responses of participants differ in some systematic way from the responses of nonparticipants. This occurs when the researcher (1) cannot locate the person (the predesignated sample element) to be studied or (2) is unsuccessful in encouraging that person to participate. This is an especially difficult problem when you are using a probability sample of subjects. Many studies have shown that better-educated individuals and those more interested in the topic participate in surveys. A high percentage of those who reply to a given survey have usually replied to others, while a large share of those who do not respond are habitual nonparticipants.<sup>6</sup>

Researchers are not without actions to avoid or diminish the error discussed above. We will explore these options in detail in Chapters 12 to 15. Despite its challenges, communicating with research participants—and the use of the survey—is the principal method of marketing research.

**Response-Based Errors** **Response error** is generated in two ways: when the participant fails to give a correct answer or fails to give the complete answer. The interviewer can do little about the participant's information level. Screening questions qualify participants when there is doubt about their ability to answer. The most appropriate applications for communication research are those where participants are uniquely qualified to provide the desired information. Questions can be used to inquire about characteristics of a participant, such as his or her household income, age, sexual preference, ethnicity, or family lifecycle stage. Questions can also be asked that reveal information exclusively internal to the participant. We include here items such as the participant's lifestyle, attitudes, opinions, expectations, knowledge, motivations, and intentions.

If we ask participants to report on events that they have not personally experienced, we need to assess the replies carefully. If our purpose is to learn what the participant understands to be the case, it is legitimate to accept the answers given. But if our intent is to learn what the event or situation actually was, we must recognize that the participant is reporting secondhand data and the accuracy of the information declines.

In the study of MindWriter's CompleteCare program, only those individuals who have experienced difficulty with their laptops and gone through the program have direct knowledge of the service process. Although some associates and family members are likely to have some secondhand knowledge of the experience, no one but the actual laptop owners is likely to give as clear a picture of what works and what doesn't with CompleteCare. The laser patient, Edna, on the other hand, had a totally different experience when she went for surgery to correct her vision. Answers to many questions on the patient survey might have been known by a caregiver, especially since Edna was experiencing eye problems serious enough to warrant surgery. And the clinic's admissions department could have been confident that such information was as accurate as it would have been if given by Edna herself. Since inaccuracy is a correctable source of error, a family or group member should not be asked about another member's experience unless there is no other way to get the information directly. We should not depend on secondhand sources if a more direct source can be found.

Participants also cause error by responding in such a way as to unconsciously or consciously misrepresent their actual behavior, attitudes, preferences, motivations, or intentions (*response bias*). Participants create response bias when they modify their responses to be socially acceptable or to save face or reputation with the interviewer (*social desirability bias*), and sometimes even in an attempt to appear rational and logical.

One major cause of response bias is *acquiescence*—the tendency to be agreeable. On the participant's part, acquiescence may be a result of lower cognitive skills or knowledge related to a concept or construct, language difficulties, or perceived level of anonymity. However, researchers can contribute to acquiescence by the speed with which they ask questions (the faster questions are asked, the more acquiescence) and the placement of questions in an interview (the later the question, the more acquiescence.)<sup>7</sup>

Sometimes participants may not have an opinion on the topic of concern. Under this circumstance, their proper response should be "don't know" or "have no opinion." Some research suggests that most participants who chose the don't-know response option actually possess the knowledge or opinion that the researcher seeks.<sup>8</sup> Participants may choose the option because they may want to shorten the time spent in the participation process, may be ambivalent or have conflicting opinions on the topic, may feel they have insufficient information to form a judgment—even though they actually have taken a position—don't believe that the response choices match their position, or don't possess the cognitive skills to understand the response options. If they choose the don't-know option for any of these reasons, studies suggest that probing for their true position will increase both reliability and validity of the data. However, forcing a participant to express some opinion he or she does not hold by withholding a don't-know option makes it difficult for researchers to know the reliability of the answers.

Participants may also interpret a question or concept differently from what was intended by the researcher. This occurs when the researcher uses words that are unfamiliar to the participant. Thus, the individual answers a question that is different from the one the researcher intended to ask. This problem is reflected in Edna's letter concerning the clinic's survey.

Regardless of the reasons, each source of participant-initiated error diminishes the value of the data collected. It is difficult for a researcher to identify such occasions. Thus, communicated responses should be accepted for what they are—statements by individuals that reflect varying degrees of truth and accuracy.

## Choosing a Communication Method

Once the sponsor or researcher has determined that surveying or interviewing is the appropriate data collection approach, various means may be used to secure information from individuals. A researcher can conduct a semistructured interview or survey by personal interview or telephone or can distribute a self-administered survey by mail, fax, computer, e-mail, the Internet, or a combination of these. As noted in Exhibit 10-5, while there are commonalities among these approaches, several considerations are unique to each.

> **Exhibit 10-5** Comparison of Communication Approaches

	Self-Administered Survey	Telephone Survey	Survey via Personal Interview
<b>Description</b>	<p>Questionnaires are:</p> <ul style="list-style-type: none"> <li>a. Mailed, faxed, or couriered to be self-administered—with return mechanism generally included (denoted below as a).</li> <li>b. Computer-delivered via intranet, Internet, and online services—computer stores/forwards completed instruments automatically (denoted below as b).</li> <li>c. People are intercepted in a central location and studied via paper or computerized instrument—without interviewer assistance; e.g., restaurant and hotel comment cards (denoted below as c).</li> </ul>	<p>People selected to be part of the sample are interviewed on the telephone by a trained interviewer.</p>	<p>People selected to be part of the sample are interviewed in person by a trained interviewer.</p>
<b>Disadvantages</b>	<ul style="list-style-type: none"> <li>• Low response rate in some modes.</li> <li>• No interviewer intervention available for probing or explanation (a).</li> <li>• Cannot be long or complex (a).</li> <li>• Accurate mailing lists needed (a).</li> <li>• Often participants returning survey represent extremes of the population—skewed responses (a).</li> <li>• Anxiety among some participants (b).</li> <li>• Directions/software instruction needed for progression through the instrument (b).</li> <li>• Computer security (b).</li> <li>• Need for low-distraction environment for survey completion (c).</li> </ul>	<ul style="list-style-type: none"> <li>• Response rate is lower than for personal interview.</li> <li>• Higher costs if interviewing geographically dispersed sample.</li> <li>• Interview length must be limited.</li> <li>• Many phone numbers are unlisted or not working, making directory listings unreliable.</li> <li>• Some target groups are not available by phone.</li> <li>• Responses may be less complete.</li> <li>• Illustrations cannot be used.</li> </ul>	<ul style="list-style-type: none"> <li>• High costs.</li> <li>• Need for highly trained interviewers.</li> <li>• Longer period needed in the field collecting data.</li> <li>• May be wide geographic dispersion.</li> <li>• Follow-up is labor-intensive.</li> <li>• Not all participants are available or accessible.</li> <li>• Some participants are unwilling to talk to strangers in their homes.</li> <li>• Some neighborhoods are difficult to visit.</li> <li>• Questions may be altered or participant coached by interviewers.</li> </ul>

In the last two decades of the 20th century, a revolution—albeit a quiet one—was under way in survey research. Driven by changing technology and the need to make research more responsible to the bottom line and ROI objectives, the paper-and-pencil survey standard of the prior 60 years was replaced by a new computerized standard. Whether it goes by the name of “computer-assisted data collection” (CADAC), “computer-assisted survey information collection” (CASIC), or “computer-assisted interviewing” (CAI), the trend is growing. While less obvious in the public sector (the U.S. government is the largest survey researcher in the world, and paper-and-pencil approaches still hold prominence there), in the private sector of survey research with households and organizations, the computer’s influence on this methodology is far-reaching. It influences all the various data collection practices.

## > Self-Administered Surveys

The **self-administered questionnaire** is ubiquitous in modern living. You have experienced service evaluations of hotels, restaurants, car dealerships, and transportation providers. Often a short questionnaire is left to be completed by the participant in a convenient location or is packaged with a product. User registrations, product information requests in magazines, warranty cards, the MindWriter CompleteCare study, and the Albany Clinic study are examples of self-administered surveys. Self-administered **mail surveys** are delivered not only by the U.S. Postal Service but also via fax and courier service. Other delivery modalities include *computer-delivered* and *intercept* studies.

## Evaluation of the Self-Administered Survey

Nowhere has the computer revolution been felt more strongly than in the area of the self-administered survey. Computer-delivered self-administered questionnaires (also labeled **computer-assisted self-interviews**, or **CASIs**) use organizational intranets, the Internet, or online services to reach their participants. Participants may be targeted (as when BizRate, an online e-business rating service, sends an e-mail to a registered e-purchaser to participate in a survey following the completion of their order) or self-selecting (as when a computer screen pop-up window offers a survey to an individual who clicks on a particular Web site or when a potential participant responds to a postcard or e-mail inquiry looking for participants). The questionnaire and its managing software may be resident on the computer or its network, or both may be sent to the participant by mail—**disk-by-mail (DBM) survey**. A 2001 Gartner Research Dataquest survey found that 61 percent of U.S. households are actively online and, once connected, 91 percent are likely to continue their Internet subscription.<sup>9</sup> Is it any wonder, then, that researchers have embraced computer-delivered self-administered surveys? See Exhibit 10-6.

Intercept surveys—at malls, conventions, state fairs, vacation destinations, even busy city street corners—may use a traditional paper-and-pencil questionnaire or a computer-delivered survey via a kiosk. The respondent participates without interviewer assistance, usually in a predetermined environment, such as a room in a shopping mall. All modes have special problems and unique advantages (as shown in Exhibit 10-5).

As computer-delivered surveys, especially those delivered via the Internet, are in their infancy, much of what researchers know about self-administered surveys has been learned from experiments conducted with mail surveys and from personal experience. So as we explore the strengths and weaknesses of the various self-administered survey methods, we will start with this body of knowledge.

> **Exhibit 10-6** The Web as a Survey Research Venue

Web Advantages	Example
Short turnaround of results: results are tallied as participants complete surveys.	A soft-drink manufacturer got results from a Web survey in just five days.
Ability to do numerous surveys over time.	A printer manufacturer did seven surveys in six months during the development of one of its latest products.
Participants feel anonymous.	Anonymity was the necessary ingredient for a study on impotence conducted by a drug manufacturer.
Experiences unavailable by other means.	<ul style="list-style-type: none"> <li>• One major advertising agency is conducting Web research using virtual supermarket aisles that participants wander through, reacting to client products and promotions.</li> <li>• LiveWorld has developed a packaging study showing more than 75 images of labels and bottle designs.</li> </ul>
Web Disadvantages (and emerging solutions)	Example
Recruiting the right sample is costly and time-consuming; unlike phone and mail sample frames, no lists exist. (Firms like Greenfield Online and Survey Samples, Inc., now provide samples built from panels of Internet users who have indicated an interest in participating in online surveys.)	Talk City, working for Wharton Associates and Fusion5, set up a panel of 3,700 teens for a survey to test new packaging for a soft drink using phone calls, referrals, e-mail lists, banner ads, and Web site visits. It drew a sample of 600 for the research. It cost more than \$50,000 to set up the list.
It takes technical as well as research skill to field a Web survey. (Numerous firms now offer survey hosting services.)	A 10- to 15-minute survey can take up to five days of technical expertise to field and test.

Source: These examples are drawn from the personal experience of the authors, as well as from Noah Shachtman, "Why the Web Works as a Market Research Tool," *AdAge.com*, Summer 2001 (<http://adage.com/tools2001>).

## &gt;snapshot

Is what you hear on the radio really the latest and greatest? Program managers say it is if you believe the research. Radio stations are now mostly owned by large chains. A single market share point shift can shift millions of dollars in revenues. So programmers increasingly rely on research to help them choose what to play. The bulk of songs are "pre-runs" that have come through song lists. Stations rely on two main types of research: play analysis and phone surveys. Play analysis tracks songs that are actually being aired on stations countrywide. Eight-second song "hooks" are recorded from more than 1,000 stations and

then cataloged by artist and title, station, and time of day. Researchers study the data for patterns, and radio chains sell the data to any station willing to pay. Phone surveys are used to test individual titles and artists. Using blocks of 25 to 30 songs, callers and on-air hosts ask listeners if the song indicates recognition, if they've heard it before, and if they'd purchase and producing companies are also willing to pay upward of \$20,000 per song to perform such tests on unknown artists. So do you hear the best and the newest or only the obnoxious best and new?

[www.clearchannel.com](http://www.clearchannel.com)

## Costs

Self-administered surveys of all types typically cost less than surveys via personal interviews. This is true of mail surveys, as well as of both computer-delivered and intercept surveys. Telephone and mail costs are in the same general range, although in specific cases either may be lower. The more geographically dispersed the sample, the more likely it is that self-administered surveys via computer or mail will be the low-cost method. A mail or computer-delivered study can cost less because it is often a one-person job. And computer-delivered studies (including those that employ interviewer-participant interaction) eliminate the cost of printing surveys, a significant cost of both mail studies and personal interviewing employing paper-and-pencil surveys. The most significant cost savings with computer-delivered surveys involve the much lower cost of pre- and postnotification (often done by mail or phone when other self-administered surveys are involved), as well as the lower per-participant survey delivery cost of very large studies.<sup>10</sup>

## Sample Accessibility

One asset to using mail self-administered surveys is that researchers can contact participants who might otherwise be inaccessible. Some groups, such as major corporate executives and physicians, are difficult to reach in person or by phone, as gatekeepers (secretaries, office managers, and assistants) limit access. But researchers can often access these special participants by mail or computer. When the researcher has no specific person to contact—say, in a study of corporations—the mail or computer-delivered survey may be routed to the appropriate participant. Additionally, the computer-delivered survey can often reach samples that are identified in no way other than their computer and Internet use, such as the users of a particular online game or those who have shopped with a particular online retailer.

## Time Constraints

While intercept studies still pressure participants for a relatively quick response, in a mail survey the participant can take more time to collect facts, talk with others, or consider replies at length than is possible in a survey employing the telephone or in a personal interview. Computer-delivered studies, especially those accessed via e-mail links to the Internet, often have time limitations on both access and completion once started. And once

started, computer-delivered studies usually cannot be interrupted by the participant to seek information not immediately known. One recent computer-delivered study sponsored by Procter & Gamble, however, asked of participants (who used skin moisturizers) the actual duration of time that the participant spent applying the product to various skin areas following a bath or shower. These questions came in the middle of a fairly lengthy survey. The participant was encouraged to discontinue the survey, time his or her moisturizer application following the next bath or shower, and return to the survey via a link and personal code with detailed responses.<sup>11</sup>

## Anonymity

Mail surveys are typically perceived as more impersonal, providing more anonymity than the other communication modes, including other methods for distributing self-administered questionnaires. Computer-delivered surveys still enjoy that same perceived anonymity, although increased concerns about privacy may erode this perception in the future.<sup>12</sup>

## Topic Coverage

A major limitation of self-administered surveys concerns the type and amount of information that can be secured. Researchers normally do not expect to obtain large amounts of information and cannot probe deeply into topics. Participants will generally refuse to cooperate with a long and/or complex mail, computer-delivered, or intercept questionnaire unless they perceive a personal benefit. Returned mail questionnaires with many questions left unanswered testify to this problem, but there are also many exceptions. One general rule of thumb is that the participant should be able to answer the questionnaire in no more than 10 minutes—similar to the guidelines proposed for telephone studies. On the other hand, one study of the general population delivered more than a 70 percent response to a questionnaire calling for 158 answers.<sup>13</sup> Several early studies of computer-delivered surveys show that participants indicate some level of enjoyment with the process, describing the surveys as interesting and amusing.<sup>14</sup> The novelty of the process, however, is expected to decline with experience, and recent declines in Web and e-mail survey response rates seem to be supporting this expectation.

## Maximizing Participation in the Self-Administered Survey

To maximize the overall probability of response, attention must be given to each point of the survey process where the response may break down.<sup>15</sup> For example:

- The wrong address, e-mail or postal, can result in nondelivery or nonreturn.
- The envelope or fax cover sheet may look like junk mail and be discarded without being opened, or the subject line on e-mail may give the impression of spam and not encourage that the e-mail be opened.
- Lack of proper instructions for completion may lead to nonresponse.
- The wrong person may open the envelope or receive the fax or e-mail and fail to call it to the attention of the right person.
- A participant may find no convincing explanation or inducement for completing the survey and thus discard it.
- A participant may temporarily set the questionnaire aside or park it in his or her e-mail in-box and fail to complete it.
- The return address may be lost, so the questionnaire cannot be returned.

Thus, efforts should be directed toward maximizing the overall probability of response. One approach, the Total Design Method (TDM), suggests minimizing the burden on participants by designing questionnaires that:<sup>16</sup>

- Are easy to read.
- Offer clear response directions.
- Include personalized communication.
- Provide information about the survey via advance notification.
- Encourage participants to respond.<sup>17</sup>

More than 200 methodological articles have been published on efforts to improve response rates. Few approaches consistently showed positive response rates.<sup>18</sup> However, several practical suggestions emerge from the conclusions:<sup>19</sup>

- Preliminary or advance notification of the delivery of a self-administered questionnaire increases response rates.
- Follow-ups or reminders after the delivery of a self-administered questionnaire increase response rates.
- Clearly specified return directions and devices (e.g., response envelopes, especially postage-stamped) improve response rates.
- Monetary incentives for participation increase response rates.
- Deadline dates do not increase response rates but do encourage participants to respond sooner.
- A promise of anonymity, while important to those who do respond, does not increase response rates.
- An appeal for participation is essential.

## Self-Administered Survey Trends

Computer surveying is surfacing at trade shows, where participants complete questionnaires while making a visit to a company's booth. Continuous tabulation of results provides a stimulus for attendees to visit a particular exhibit as well as gives the exhibitor detailed information for evaluating the productivity of the show. This same technology easily transfers to other situations where large groups of people congregate.

Companies are now using intranet capabilities to evaluate employee policies and behavior. Ease of access to electronic mail systems makes it possible for both large and small organizations to use computer surveys with both internal and external participant groups. Many techniques of traditional mail surveys can be easily adapted to computer-delivered questionnaires (e.g., follow-ups to nonparticipants are more easily executed and are less expensive).

It is not unusual to find registration procedures and full-scale surveying being done on World Wide Web sites. University sites are asking prospective students about their interests, and university departments are evaluating current students' use of online materials. A short voyage on the Internet reveals organizations using their sites to evaluate customer service processes, build sales-lead lists, evaluate planned promotions and product changes, determine supplier and customer needs, discover interest in job openings, evaluate employee attitudes, and more. Advanced and easier-to-use software for designing Web questionnaires is no longer a promise of the future—it's here.

*You will find a useful Web questionnaire tutorial at <http://www.surveypro.com/>.*

**The Web-based questionnaire**, a measurement instrument both delivered and collected via the Internet, has the power of computer-assisted telephone interview systems, but without the expense of network administrators, specialized software, or additional hardware. As a solution for Internet or intranet Web sites, you need only a personal computer and Web access. Most software products are wizard driven with design features that allow custom survey creation and modification.



Two primary options are proprietary solutions offered through research firms and off-the-shelf software designed for researchers who possess the knowledge and skills we describe here and in Chapter 14. With fee-based services, you are guided (often online) through problem formulation, questionnaire design, question content, response strategy, and wording and sequence of questions. The supplier's staff generates the questionnaire HTML code, hosts the survey at their server, and provides data consolidation and reports. Off-the-shelf software is a strong alternative. *PC Magazine* reviewed six packages containing well-designed user interfaces and advanced data preparation features.<sup>20</sup> The advantages of these software programs are:

- Questionnaire design in a word processing environment.
- Ability to import questionnaire forms from text files.
- A coaching device to guide you through question and response formatting.
- Question and scale libraries.
- Automated publishing to a Web server.
- Real-time viewing of incoming data.
- Ability to edit data in a spreadsheet-type environment.
- Rapid transmission of results.
- Flexible analysis and reporting mechanisms.

Ease of use is not the only influence pushing the popularity of Web-based instruments. Cost is a major factor. A Web survey is much less expensive than conventional survey research. Although fees are based on the number of completions, the cost of a sample of 100 might be one-sixth that of a conventional telephone interview. Bulk mailing and e-mail data collection have also become more cost-effective because any instrument may be configured as an e-mail questionnaire.

The computer-delivered survey has made it possible to use many of the suggestions for increasing participation. Once the computer-delivered survey is crafted, the cost of redelivery via computer is very low. Preliminary notification via e-mail is both more timely and less costly than notification for surveys done by phone or mail. The click of a mouse or a single keystroke returns a computer-delivered study. Many computer-delivered surveys use color, even color photographs, within the survey structure. This is not a cost-effective option with paper surveys. And video clips—never an option with a mail survey—are possible with a computer-delivered survey. In addition, e-currencies have simplified the delivery of monetary and other incentives. However, employing all the stimulants for participation cannot overcome technology snafus. These glitches are likely to continue to plague participation as long as researchers and participants use different computer platforms, operating systems, and software.

While Web- and e-mail-based self-administered surveys have certainly caught the lion's share of business attention in the last few years, the tried-and-true methods of telephone and personal interviews still have their strengths—and their advocates in the research community.

## > Survey via Telephone Interview

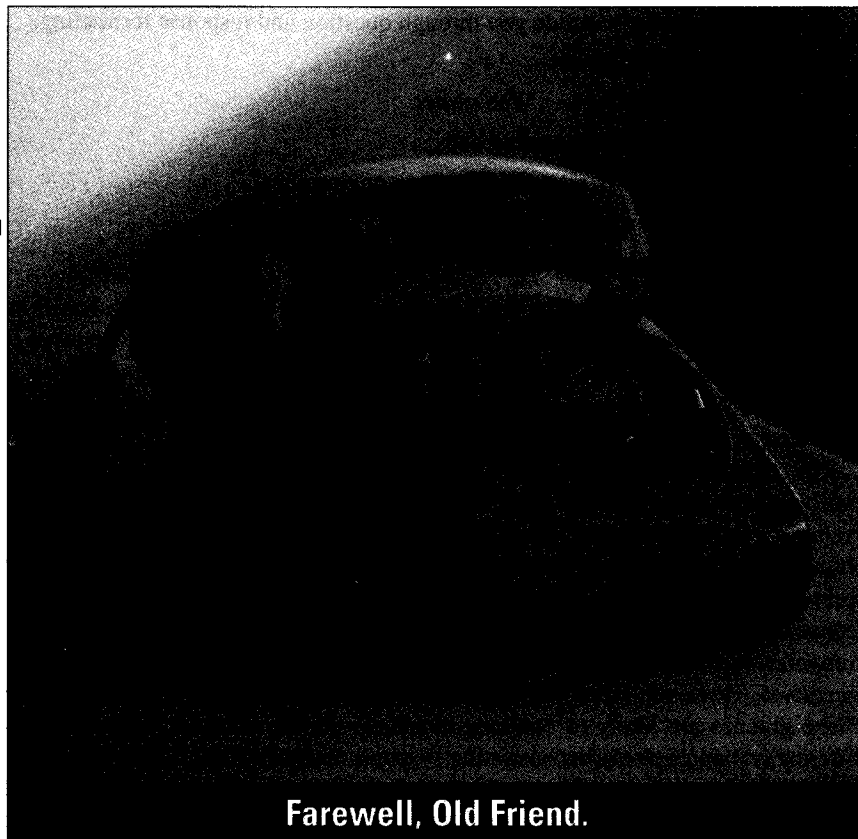
The **telephone survey** is still the workhorse of survey research. With the high level of telephone service penetration in the United States and the European Union, access to participants through low-cost, efficient means has made telephone interviewing a very attractive alternative for researchers. Nielsen Media Research uses thousands of calls each week to determine television viewing habits, and Arbitron does the same for radio listening habits. Pollsters working with political candidates use telephone surveys to assess the power of a speech or a debate during a hotly contested campaign. Numerous firms field phone omnibus studies each week. Individual questions in these studies are used to capture everything from people's feeling about the rise

in gasoline prices to the power of a celebrity spokesperson in an advertising campaign or the latest teenage fashion trend.

## Evaluation of the Telephone Interview

Of the advantages that telephone interviewing offers, probably none ranks higher than its moderate cost. One study reports that sampling and data collection costs for telephone surveys can run from 45 to 64 percent lower than costs for comparable personal interviews.<sup>21</sup> Much of the savings comes from cuts in travel costs and administrative savings from training and supervision. When calls are made from a single location, the

In the September 4, 2003, Harris Poll®, 42 percent of adults erroneously thought registering for the national Do Not Call registry would ban telephone survey calls as well. While researchers are aggressively exploring online research, the phone survey still plays an important role in business research.




**Farewell, Old Friend.**

There are now more than 50 million consumer phone numbers in the National Do Not Call Registry – and it's growing every day. To complicate matters, almost half\* of all Americans incorrectly believe that telephone survey calls are now banned by the Do Not Call rules.

The future of telephone survey work is in doubt. But there is no doubt that viable, accurate data collection alternatives exist.

Call Harris Interactive today to learn more – we'll be happy to answer the phone.

 **HarrisInteractive**  
MARKET RESEARCH

\*42% of all U.S. adults think that the registry applies to calls to conduct surveys about products and services – "The Survey Fall" – September 4, 2003.

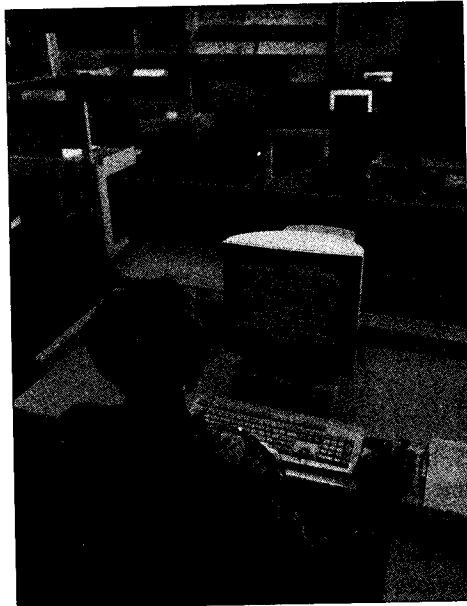
[www.harrisinteractive.com/DNC](http://www.harrisinteractive.com/DNC) Tel 877.919.4765

©2005 Harris Interactive Inc. All rights reserved. 001 01/2/05 DL-05

researcher may use fewer, yet more skilled, interviewers. Telephones are especially economical when callbacks to maintain precise sampling requirements are necessary and participants are widely scattered. Long-distance service options make it possible to interview nationally at a reasonable cost.

Telephone interviewing can be combined with immediate entry of the responses into a data file by means of terminals, personal computers, or voice data entry. This brings added savings in time and money. The **computer-assisted telephone interview (CATI)** is used in research organizations throughout the world. A CATI facility consists of acoustically isolated interviewing carrels organized around supervisory stations. The telephone interviewer in each carrel has a personal computer or terminal that is networked to the phone system and to the central data processing unit. A software program that prompts the interviewer with introductory statements, qualifying questions, and precoded questionnaire items drives the survey. These materials appear on the interviewer's monitor. CATI works with a telephone number management system to select numbers, dial the sample, and enter responses. One facility, the Survey Research Center at the University of Michigan, consists of 60 carrels with 100 interviewers working in shifts from 8 a.m. to midnight (EST) to call nationwide. When fully staffed, it produces more than 10,000 interview hours per month.<sup>22</sup>

Another means of securing immediate response data is the **computer-administered telephone survey**. Unlike CATI, there is no human interviewer. A computer calls the phone number, conducts the interview, places data into a file for later tabulation, and terminates the contact. The questions are voice-synthesized, and the participant's answers and computer timing trigger continuation or disconnect. Several modes of computer-administered surveys exist, including *touch-tone data entry (TDE)*; *voice recognition (VR)*, which recognizes a limited vocabulary, usually yes/no responses; and *automatic speech recognition (ASR)* for



>picprofile

RTI International Call Center Services employs over 400 interviewers, institutional contacts, quality control monitors, team leaders, and supervisors in its two state-of-the-art call centers located in Charlotte and Raleigh, North Carolina. The call centers typically conduct between 10 and 30 different data collection projects each month, with staff members completing more than 100,000 telephone interviews annually. Call center staff come from all backgrounds, many are students or others who work as interviewers part-time in the evenings and on weekends. RTI International conducts rigorous training with all telephone staff on standardized interviewing techniques, including the use of the telephone, rapport cooperation, and the use of its computer-assisted telephone interviewing system.

recognizing and recording a wide range of verbal responses. CATI is often compared to the self-administered questionnaire and offers the advantage of enhanced participant privacy. One study showed that the noncontact rate for this electronic survey mode is similar to that for other telephone interviews when a random phone list is used. It also found that rejection of this mode of data collection affects the refusal rate (and thus non-response error) because people hang up more easily on a computer than on a human.<sup>23</sup> The **noncontact rate** is a ratio of potential but unreached contacts (no answer, busy, answering machine or voice mail, and disconnects but not refusals) to all potential contacts.

The **refusal rate** refers to the ratio of contacted participants who decline the interview to all potential contacts. New technology, notably call-filtering systems where the receiver can decide whether a call is answered based on caller identity, is expected to increase the noncontact rate associated with telephone surveys. The 2003 CMOR Respondent Cooperation and Industry Image Study reported that while survey refusal rates have been growing steadily over several years, the rate “took a sharper than usual increase” this year. The study also noted that “positive attitudes [about participating in surveys] are declining, while negative perceptions are increasing.”<sup>24</sup>

When compared to either personal interviews or mail self-administered surveys, the use of telephones brings a faster completion of a study, sometimes taking only a day or so for the fieldwork. When compared to personal interviewing, it is also likely that interviewer bias, especially bias caused by the physical appearance, body language, and actions of the interviewer, is reduced by using telephones.

Finally, behavioral norms work to the advantage of telephone interviewing. If someone is present, a ringing phone is usually answered, and it is the caller who decides the purpose, length, and termination of the call.<sup>25</sup>

There are also disadvantages to using the telephone for research. A skilled researcher will evaluate the use of a telephone survey to minimize the effect of these disadvantages:

- Inaccessible households (no telephone service or no/low contact rate).
- Inaccurate or nonfunctioning numbers.
- Limitation on interview length (fewer measurement questions).
- Limitations on use of visual or complex questions.
- Ease of interview termination.
- Less participant involvement.
- Distracting physical environment.

## Inaccessible Households

Approximately 94 percent of all U.S. households have access to telephone service.<sup>26</sup> On the surface, this should make telephone surveys a prime methodology for communication studies. However, several factors reduce such an enthusiastic embrace of the methodology. Rural households and households with incomes below the poverty line remain underrepresented in telephone studies, with phone access below 75 percent.<sup>27</sup> More households are using filtering devices and services to restrict access, including caller ID, privacy manager, Tele-Zapper, and unlisted numbers (estimated between 22 and 30 percent of all household phone numbers).<sup>28</sup> Meanwhile, the number of inaccessible individuals continues to increase as cellular/wireless phone use increases. From 1985 to 2002, the number of U.S. wireless telecommunication subscribers grew from 203.6 thousand to 134.5 million.<sup>29</sup> Many of these numbers are unlisted or possess screening or filtering services. Additionally, people’s use of phone modems to access the Internet makes household lines ring busy for long periods of time. Recent FCC filings indicate that fewer than 15 percent of U.S. households have second telephone lines, required for simultaneous Internet access.<sup>30</sup> Effective May 2004 federal wireless local-number portability legislation made it possible for subscribers to take their wired phone number to their wireless

phone service (or the reverse) or to shift their wireless service between carriers without losing their wireless number. Thus the guidelines for identifying the physical location of a phone by its number—and, in turn, the location of its owner—no longer apply.<sup>31</sup>

These causes of variations in participant availability by phone can be a source of bias. A random dialing procedure is designed to reduce some of this bias. **Random dialing** normally requires choosing phone exchanges or exchange blocks and then generating random numbers within these blocks for calling.<sup>32</sup> Of course, just reaching a household doesn't guarantee its participation.

## Inaccurate or Nonfunctioning Numbers

One source says the highest incidence of unlisted numbers is in the West, in large metropolitan areas, among nonwhites, and for persons between 18 and 34 years of age.<sup>33</sup> Several methods have been developed to overcome the deficiencies of directories; among them are techniques for choosing phone numbers by using random dialing or combinations of directories and random dialing.<sup>34</sup> However, increasing demand for multiple phone lines by both households and individuals has generated new phone area codes and local exchanges. This too increases the inaccuracy rate.

## Limitation on Interview Length

A limit on interview length is another disadvantage of the telephone survey, but the degree of this limitation depends on the participant's interest in the topic. Ten minutes has generally been thought of as ideal, but interviews of 20 minutes or more are not uncommon. One telephone survey sponsored by Kraft lasted approximately 30 minutes. It was designed to judge the willingness of sample issue recipients to subscribe to a prototype magazine, *food&family*. The survey also measured the effectiveness of the sample issue of the magazine to deliver purchase intent for Kraft products featured in the recipes contained therein.<sup>35</sup> In another study, interviews ran for one and a half hours in a survey of long-distance services.<sup>36</sup>

## Limitations on Use of Visual or Complex Questions

The telephone survey limits the complexity of the survey and the use of complex scales or measurement techniques that is possible with personal interviewing, CASI, or WWW surveys. For example, in personal interviews, participants are sometimes asked to sort or rank an array of cards containing different responses to a question. For participants who cannot visualize a scale or other measurement device that the interview is attempting to describe, one solution has been to employ a nine-point scaling approach and to ask the participant to visualize it by using the telephone dial or keypad.<sup>37</sup> In telephone interviewing it is difficult to use maps, illustrations, and other visual aids. In some instances, however, interviewers have supplied these prior to a prescheduled interview via fax, e-mail, or the Internet.

## Ease of Interview Termination

Some studies suggest that the response rate in telephone studies is lower than that for comparable face-to-face interviews. One reason is that participants find it easier to terminate a phone interview. Telemarketing practices may also contribute. Public reaction to investigative reports of wrongdoing and unethical behavior within telemarketing activities places an added burden on the researcher, who must try to convince a participant that the phone interview is not a pretext for soliciting contributions (labeled *frugging*—fund-raising under the guise of research) or selling products (labeled *sugging*—sales under the guise of research).

## Less Participant Involvement

Telephone surveys can result in less thorough responses, and persons interviewed by phone find the experience to be less rewarding than a personal interview. Participants report less rapport with telephone interviewers than with personal interviewers. Given the growing costs and difficulties of personal interviews, it is likely that an even higher share of surveys will be by telephone in the future. Thus, it behooves researchers using telephone surveys to attempt to improve the enjoyment of the interview. One authority suggests:

We need to experiment with techniques to improve the enjoyment of the interview by the participant, maximize the overall completion rate, and minimize response error on specific measures. This work might fruitfully begin with efforts at translating into verbal messages the visual cues that fill the interaction in a face-to-face interview: the smiles, frowns, raising of eyebrows, eye contact, etc. All of these cues have informational content and are important parts of the personal interview setting. We can perhaps purposefully choose those cues that are most important to data quality and participant trust and discard the many that are extraneous to the survey interaction.<sup>38</sup>

## Changes in the Physical Environment

Replacement of home or office phones with cellular and wireless phones also raises concerns. In regard to telephone surveys, researchers are concerned about the changing environment in which such surveys might be conducted, the resulting quality of data collected under possibly distracting circumstances—at a busy intersection, in the midst of weekly shopping in a congested grocery aisle, at the local high school basketball tournament—and the possible increase in refusal rates.

## Telephone Survey Trends

Future trends in telephone surveying bear watching. Answering machines or voice-mail services pose potentially complex response rate problems since they are estimated to have substantial penetration in American households. Previous research discovered that most such households are accessible; the subsequent contact rate was greater in answering-machine households than in no-machine households and about equal with busy-signal households. Other findings suggested that (1) individuals with answering machines were more likely to participate, (2) machine use was more prevalent on weekends than on weekday evenings, and (3) machines were more commonplace in urban than in rural areas.

## >snapshot

### A Zap-Attack on Telephone Surveys

Privacy concerns have been fueling a decline in telephone survey participation for the last decade, but a product introduced late in 2001 may have even further-reaching ramifications for this type of research. The Tele-Zapper, by Privacy Technologies, Inc., detects predictive-dialer computers' incoming calls, disconnects such a call before contact is made between the caller and the potential participant, and then sends a signal to the dialer that the phone number is disconnected. When the predictive dialer hears the signal tone, it disconnects and removes the number from the call list. While the device was intended to "zap" unwanted calls from telemarketers, Privacy Technologies claims that calls from other

organizations that use predictive-dialer computers (charitable organizations, opinion and political pollsters, as well as marketing researchers) also will be zapped.

One saving grace for researchers is the \$49.95 price. To many, the price may outweigh the inconvenience of the disruptive call. But to others, the price may seem a small investment in protection against marketing "scams." Another bright spot for researchers is that not all use predictive dialers—manually dialed calls to a potential research participant cannot be zapped.

[www.telezapper.com](http://www.telezapper.com)

Voice-mail options offered by local phone service providers have less market penetration but are gaining increasing acceptance. Questions about the sociodemographics of users and nonusers and the relationship of answering-machine/voice-mail technology to the rapid changes in the wireless market remain to be answered.<sup>39</sup> Caller identification technology, the assignment of facsimile machines or computer modems to dedicated phone lines, and technology that identifies computer-automated dialers and sends a disconnect signal in response are all expected to have an impact on the noncontact rate of phone interviews.

The variations among the 60 telephone companies' services and the degree of cooperation that will be extended to researchers are also likely to affect noncontact rates. There is also concern about the ways in which random dialing can be made to deal with nonworking and ineligible numbers.<sup>40</sup> But arguably no single threat poses greater danger than the government-facilitated Do Not Call registry initiated in 2003 by the Federal Trade Commission. More than 50 million U.S. household and cell numbers, a third of all U.S. households, were registered in its initial wave of enrollment.<sup>41</sup> While currently survey researchers are exempt from its restrictions, customer confusion about the distinction between research and telemarketing is likely to cause an increase in the nonresponse rate. Telemarketers might be the catalyst, but legitimate research will suffer.

## > Survey via Personal Interview

A **survey via personal interview** is a two-way conversation between a trained interviewer and a participant. With her poor eyesight and the problems of question clarity, a personal interview, rather than the intercept/self-administered questionnaire, might have been a preferable communication method for Edna at the Albany Outpatient Laser Clinic.

### Evaluation of the Personal Interview Survey

< We discuss the individual depth interview at length in Chapter 8.

There are real advantages as well as clear limitations to surveys via personal interview. The greatest value lies in the depth of information and detail that can be secured.

## >snapshot

### Aleve: Personal Interviews Provide Relief

Bayer Consumer Care inherited Aleve, a long-duration over-the-counter painkiller, from Procter & Gamble in 1996. Since its launch in 1994, P&G hadn't been able to move Aleve beyond a 6 percent market share. Bayer chose CLT Research Associates to identify potential Aleve users. CLT conducted in-home interviews with a random sample of 800 men and women aged 18 to 75 who had used a nonprescription pain reliever in the past year. The research revealed that 24 percent of those interviewed could be defined as "pain-busters" (heavy users of analgesics who were likely to try new products to gain relief). More than one-third of those identified as pain-busters had tried Aleve. Bayer's task was to use the research to identify a strategy to get pain-busters to choose Aleve when they faced their analgesic-stocked medicine cabinet.

First, Moskowitz Jacobs, Inc., had 249 participants rate various statements about Aleve. Statements that promised

"control over pain" or "freedom to do the things you want" were discovered as important emotional triggers for consumers interested in minimizing the number of pills they took to relieve pain—Aleve's differential benefit. Next Bayer managers analyzed syndicated data from Medioscope, Nielsen Panel Data, MRI, and Simmons and conducted a series of focus groups moderated by Viewpoints Consulting, Inc., to flesh out findings. The sum of this research revealed Aleve users were more likely to suffer from arthritis and back pain than the average analgesic user. This helped Bayer define the benefit of Aleve as "liberation from tough pain, making a dramatic difference in the quality of life." The resulting Dramatic Difference ad campaign boosted the subsequent year's sales by 16 percent, with a rise in market share to 7 percent, its highest ever.

[www.aleve.com](http://www.aleve.com); [www.cltresearch.com](http://www.cltresearch.com); [www.smr.com](http://www.smr.com); <http://acnlisen.com>

In the best-seller *Permission Marketing: Turning Strangers into Friends, and Friends into Customers*, Seth Godin and Don Peppers focused marketers' attention on two business practices, which they defined as interruption marketing and permission marketing. In the most popular e-book to date, *Unleashing the Idiot*, Godin discusses the power of amplified word of mouth. Both these manifests offer lessons for researchers.

#### INTERRUPTION MARKETING

Godin claims that interruption marketing is what most marketers do and have been doing for centuries. They yell out their benefits as we pass by their exhibit; they interrupt our dinner with their telemarketing calls; they interrupt our movie, radio, or television shows with commercials; they interrupt our stock news with internet pop-up ads. Interruption marketing needs to rely increasingly on creativity—or even shock—to grab our attention. The more marketers do this, the more intrusive to interruption we become. Godin claims that the interruption model worked as long as there was not an overabundance of interruptions. But the model isn't working as well as it used to in an environment where we are bombarded with more than 3,000 marketing messages a day. The problem? As disoriented as you and I are going to pay to watch the Super Bowl, you will not pay to deliver an advertisement to the Super Bowl from the Super Bowl even though you could have watched it on commercial-loaded TV or made a purchased form of the *Play or Mania* so that you could watch it again and again without interruption.

#### PERMISSION MARKETING

Permission marketing operates on a different premise. The marketer gives its customer the right, the encouragement, and the power to engage in a conversation with the company, as well as with other prospects and customers. The customer decides whether to maintain the conversation or discontinue it. Customers give permission for intrusion because they perceive a reward. If the marketer fails to deliver the reward, the customer will withdraw the permission—the first step in severing a relationship. If a potential customer drops a hardball and asks a sales person entrance, the marketer has effectively been told, "I'm not interested." The marketer can decide to send a barrage of pay messages at that disengaged individual, but the more it interrupts, the more it is likely to gain an enquiring enemy rather than a partner. Godin considers e-mail the ultimate conversational medium and a powerful one when married with internet content. "If you get permission to use e-mail to deliver marketing messages, and if people agree to participate in the messages, you've changed the game." The power of permission is that the marketer is not faced with the threat of anybody ever being able to initiate a conversation with a prospect or customer.

In one *FastCompany* article, Godin noted three major anticipated shifts in marketing's future: (1) that conversations between a marketer and its customers would become increasingly less viable to competitors and more viable if the marketer established a bond between marketer and customer; (2) that the bond between marketer and customer would have to become increasingly virginal, and (3) that conversations would increasingly substitute to the information they thought had gone. In order to avoid interruptions they thought had gone.

It far exceeds the information secured from telephone and self-administered studies via mail or computer (both intranet and Internet). The interviewer can also do more things to improve the quality of the information received than is possible with another method.

The absence of assistance in interpreting questions in the Albany Clinic study was a clear weakness that would have been improved by the presence of an interviewer. Interviewers can note conditions of the interview, probe with additional questions, and gather supplemental information through observation. Edna was obviously in good spirits and very relaxed after she and her fellow patients had critiqued the questionnaire. This attitude would have been observed and noted by an interviewer. Of course, we're hopeful that the interviewer would correctly interpret laughter as a sign of humor and not as a negative attitude, as did the admissions clerk.

Human interviewers also have more control than other kinds of communication studies. They can pre-screen to ensure the correct participant is replying, and they can set up and control interviewing conditions. They can use special scoring devices and visual materials, as is done with a **computer-assisted personal interview (CAPI)**. Interviewers also can adjust the language of the interview as they observe the problems and effects the interview is having on the participant.



**PERMISSION RESEARCH**

While Godin's book is focused on what advertisers have to do to draw our attention, there are valuable lessons for survey researchers as well. The logical research extension of Godin's argument is that marketers should offer customers or receptive prospects the opportunity to participate in survey research. This is exactly what several companies are doing. Greenfield Online has built an e-mail database of individuals who are willing to participate in surveys for a large and varied group of companies. Procter & Gamble has amassed a permission survey population to test new concepts and anticipate the effects of attitude and lifestyle changes on current products and their marketing. BusinessWeek invited its subscriber group to join a survey panel, which it later queried about the attractiveness of a possible new venture, BusinessWeek, the television show. And such surveys don't have to be delivered via e-mail or the Internet, although increasingly they are. Once the participant has given permission for the research company to seek involvement, this is an indication that the survey—or the panel drive—can be delivered by a variety of media.

As retail rates rise, permission surveying—the act of turning prospects or customers who have given permission to such engagement—should become increasingly attractive. One recent study reported in the *Journal of Advertising Research* noted that, contrary to an issue in the sampling population has strong positive correlation with response rates, regardless of the survey delivery method.<sup>43</sup> Marketing decisions are increasingly complex in an environment of product proliferation and lack of true product distinction. The survey remains the researcher's process for collecting large amounts of detailed information from willing participants. So what if we could turn a willing participant into an enthusiastic one? Using Godin's approach, survey re-

searchers should be able to marry the promise of mutual benefit and participant enthusiasm to ensure that participants in longer surveys or more complex activities that are more likely to reveal true motivations or biased attitudes.

Discontinuation/discontinuity are also major concerns in survey research. Participants discontinue surveys when their life interrupts. Discontinuation would be reduced and surveys should be returned more quickly in a permission surveying population.

Most marketers would think that their survey activity be invisible to their competitor. Competitors can use survey activity to reveal what their firm is considering. In some instances those same marketers can use their new-found knowledge to preempt a competitive move. One recent permission survey explored a new product concept for a dairy product that has visible cosmetic effects: radiant, even skin tone and clear hair. A survey was used to determine whether the promise of the new product was believable and embraceable or would be rejected. The more surveys are used to gain understanding that is applied to making strategic and tactical decisions, the more attractive the possibility of permission surveying will become.

Godin believes that the core test on the brand-owning end and that what free doctors build after the marketing impact shake-out will be the key to get an idea that qualified word of mouth started. He's convinced that marketers who use customers as "strangers" (those who gain ideas for sharing valuable new ideas with others) will be the true beneficiaries of an idea virus, and those customers will act in for other types of behaviors that give them ideas to spread. Participating in a survey is but one more way marketers can enhance their position. They will know things that others don't or know them long before most others will.

[www.sathgodin.com](http://www.sathgodin.com)

With such advantages, why would anyone want to use any other survey method? Probably the greatest reason is that personal interviewing is costly, in terms of both money and time. A survey via personal interview may cost anywhere from a few dollars to several hundred dollars for an interview with a hard-to-reach person. Costs are particularly high if the study covers a wide geographic area or has stringent sampling requirements. An exception to this is the survey via **intercept interview** that targets participants in centralized locations such as retail malls or, as with Edna, in a doctor's office. Intercept interviews reduce costs associated with the need for several interviewers, training, and travel. Product and service demonstrations also can be coordinated, further reducing costs. Their cost-effectiveness, however, is offset when representative sampling is crucial to the study's outcome. The intercept survey would have been a possibility in the Albany Clinic study, although more admissions clerks would likely have been needed if volunteers were not available to perform this task.

Costs have risen rapidly in recent years for most communication methods because changes in the social climate have made personal interviewing more difficult. Many people today are reluctant to talk with strangers or to permit strangers to visit in their homes. Interviewers are reluctant to visit unfamiliar

*You will find tips on intercept surveys on the text CD.*

neighborhoods alone, especially for evening interviewing. Finally, results of surveys via personal interviews can be affected adversely by interviewers who alter the questions asked or in other ways bias the results. As Edna and her friends discussed the Albany Clinic survey, they each applied their own operational definitions to the concepts and constructs being asked. This confusion created a bias that might have been eliminated by a well-trained interviewer. Interviewer bias, identified as one of the three major sources of error in Exhibit 10-3, was discussed earlier in this chapter. If we are to overcome these deficiencies, we must appreciate the conditions necessary for interview success.

## > Selecting an Optimal Survey Method

The choice of a communication method is not as complicated as it might first appear. By comparing your research objectives with the strengths and weaknesses of each method, you will be able to choose one that is suited to your needs. The summary of advantages and disadvantages of personal interviews, telephone interviews, and self-administered questionnaires presented in Exhibit 10-5 should be useful in making such a comparison.

When your investigative questions call for information from hard-to-reach or inaccessible participants, the telephone interview, mail survey, or computer-delivered survey should be considered. However, if data must be collected very quickly, the mail survey would likely be ruled out because of lack of control over the returns. Alternatively, you may decide your objective requires extensive questioning and probing; then the survey via personal interview should be considered.

If none of the choices turns out to be a particularly good fit, it is possible to combine the best characteristics of two or more alternatives into a *hybrid* survey. Although this decision will incur the costs of the combined modes, the flexibility of tailoring a method to your unique needs is often an acceptable trade-off.

In the MindWriter study, Jason Henry plans to insert a postcard questionnaire (a self-administered survey delivered via courier) in each laptop returned by the CompleteCare repair service. But this plan is not without problems. Not all customers will return their questionnaires, creating nonresponse bias. The postcard format doesn't permit much space for encouraging customer response. Alerting customers to the importance of returning the response card by phone (to announce courier delivery of a repaired laptop) might improve the research design, but it would be too costly when 10,000 units are processed monthly. Participants would not be in the best frame of mind if they received a damaged laptop; dissatisfaction could lead to a decreased response rate and an increase in call center contacts. Jason's proposal contains a follow-up procedure—telephoning nonparticipants to obtain their answers when response cards are not returned. This will likely decrease nonresponse error. Where most of the study participants are answering measurement questions without assistance, telephone interviewing creates the possibility of interviewer bias at an unknown level for at least part of the data.

In the Albany Clinic study, the researcher could have taken several actions to improve the quality of the data. Distributing the questionnaire to the patient's eye doctor or to the patient (by mail) prior to arrival would have increased the accuracy of identifying medications, diagnoses, hospitalizations, and so forth. The patient's eye doctor was in the best position to encourage compliance with the collection process but was not consulted. Having the patient bring the completed questionnaire to the admissions procedure, where the admissions clerk could review the completed instrument for accuracy and completeness, would have given the researcher the opportunity to clarify any confusion with the questions, concepts, and constructs. Finally, pretesting the instrument with a sample of patients would have revealed difficulties with the process and operational definitions. Edna's concerns could have been eliminated before they surfaced.

Ultimately, all researchers are confronted by the practical realities of cost and deadlines. As Exhibit 10-5 suggests, on the average, surveys via personal interview are the most expensive communication method and

take the most field time unless a large field team is used. Telephone surveys are moderate in cost and offer the quickest option, especially when CATI is used. Questionnaires administered by e-mail or the Internet are the least expensive. When your desired sample is available via the Internet, the Internet survey may prove to be the least expensive communication method with the most rapid (simultaneous) data availability. The use of the computer to select participants and reduce coding and processing time will continue to improve the cost-to-performance profiles of this method in the future.

Most of the time, an optimal method will be apparent. However, managers' needs for information often exceed their internal resources. Such factors as specialized expertise, a large field team, unique facilities, or a rapid turnaround prompt managers to seek assistance from research vendors of survey-related services.

## Outsourcing Survey Services

Commercial suppliers of research services vary from full-service operations to specialty consultants. When confidentiality is likely to affect competitive advantage, the manager or staff will sometimes prefer to bid only a phase of the project. Alternatively, the organization's staff members may possess such unique knowledge of a product or service that they must fulfill a part of the study themselves. Regardless, the exploratory work, design, sampling, data collection, or processing and analysis may be contracted separately or as a whole. Most organizations use a request for proposal (RFP) to describe their requirements and seek competitive bids (see the sample RFP in Chapter 4).

< **Types of research suppliers were discussed in Chapter 1.**

Research firms also offer special advantages that their clients do not typically maintain in-house. Centralized-location interviewing or computer-assisted telephone facilities may be particularly desirable for certain research needs. A professionally trained staff with considerable experience in similar management problems is another benefit. Data processing and statistical analysis capabilities are especially important for some projects. Other vendors have specially designed software for interviewing and data tabulation.<sup>44</sup> Panel suppliers provide another type of research service, with emphasis on longitudinal survey work.<sup>45</sup> By using the same participants over time, a **panel** can track trends in attitudes toward issues or products, product adoption or consumption behavior, and a myriad of other research interests. Suppliers of panel data can secure information from personal and telephone interviewing techniques as well as from the mail, the Web, and mixed-modes surveys. Diaries are a common means of chronicling events of research interest by the panel members. These are mailed back to the research organization. Point-of-sale terminals and scanners aid electronic data collection for panel-type participant groups. And mechanical devices placed in the homes of panel members may be used to evaluate media usage. ACNielsen, Yankelovich Partners, The Gallup Organization, and Harris Interactive all manage extensive panels.

## >summary

- 1 The communication approach involves surveying or interviewing people and recording their responses for analysis. Communication is accomplished via personal interviews, telephone interviews, or self-administered surveys, with each method having its specific strengths and weaknesses. The optimal communication method is the one that is instrumental for answering your research question and dealing with the constraints imposed by time, budget, and human resources. The opportunity to combine several survey methodologies makes the use of the mixed mode desirable in many projects.
- 2 Successful communication requires that we seek information the participant can provide and that the participant understand his or her role and be motivated to play that role. Motivation, in particular, is a task for the interviewer. Good rapport with the participant should be established quickly, and then the technical process of collecting data should begin. The latter often calls for skillful probing to supplement the answers volunteered by

the participant. Simplicity of directions and instrument appearance are additional factors to consider in encouraging response in self-administered communication studies.

- 3 Two factors can cause bias in interviewing. One is non-response. It is a concern with all surveys. Some studies show that the first contact often secures less than 20 percent of the designated participants. Various methods are useful for increasing this representation, the most effective being making callbacks until an adequate number of completed interviews have been secured. The second factor is response error, which occurs when the participant fails to give a correct or complete answer. The interviewer also can contribute to response error. The interviewer can provide the main solution for both of these two types of errors.
- 4 The self-administered questionnaire can be delivered by the U.S. Postal Service, facsimile, a courier service, a computer, or an intercept. Computer-delivered self-administered questionnaires use organizational intranets, the Internet, or online services to reach their participants. Participants may be targeted or self-selecting. Intercept studies may use a traditional questionnaire or a computerized instrument in environments where interviewer assistance is minimal.

Telephone interviewing remains popular because of the diffusion of telephone service in households and the low cost of this method compared with personal interviewing. Long-distance telephone interviewing has grown. There are also disadvantages to telephone interviewing. Many phone numbers are unlisted, and directory listings become obsolete quickly. There is also a limit on the length and depth of interviews conducted using the telephone.

The major advantages of personal interviewing are the ability to explore topics in great depth, achieve a high degree of interviewer control, and provide maximum interviewer flexibility for meeting unique situations. However, this method is costly and time-consuming, and its flexibility can result in excessive interviewer bias.

- 5 Outsourcing survey services offers special advantages to managers. A professionally trained research staff, centralized-location interviewing, focus group facilities, and computer-assisted facilities are among them. Specialty firms offer software and computer-based assistance for telephone and personal interviewing as well as for mail and mixed modes. Panel suppliers produce data for longitudinal studies of all varieties.

## >keyterms

communication approach 245

computer-administered  
telephone survey 261

computer-assisted personal  
interview (CAPI) 266

computer-assisted self-interview  
(CASI) 254

computer-assisted telephone  
interview (CATI) 261

disk-by-mail (DBM) survey 254

intercept interview 267

interviewer error 246

mail survey 254

noncontact rate 262

nonresponse error 251

panel 269

random dialing 263

refusal rate 262

response error 251

self-administered survey 254  
survey 245

survey via personal interview 265

telephone survey 259

Web-based questionnaire 258

## >discussionquestions

### Terms in Review

- 1 Distinguish among response error, interviewer error, and nonresponse error.
- 2 How do environmental factors affect response rates in personal interviews? How can we overcome these environmental problems?

### Making Research Decisions

- 3 Assume you are planning to interview shoppers in a shopping mall about their views on increased food prices and what the federal government should do about them. In what different ways might you try to motivate shoppers to cooperate in your survey?
  - 4 In recent years, in-home personal interviews have grown more costly and more difficult to complete. Suppose, however, you have a project in which you need to talk with people in their homes. What might you do to hold down costs and increase the response rate?
  - 5 In the following situations, decide whether you would use a personal interview, telephone survey, or self-administered questionnaire. Give your reasons.
    - a A survey of the residents of a new subdivision on why they happened to select that area in which to live. You also wish to secure some information about what they like and do not like about life in the subdivision.

- b A poll of students at Metro University on their preferences among three candidates who are running for president of the student government.
  - c A survey of 58 wholesale grocery companies, scattered over the eastern United States, on their personnel management policies for warehouse personnel.
  - d A survey of financial officers of the Fortune 500 corporations to learn their predictions for the economic outlook in their industries in the next year.
  - e A study of applicant requirements, job tasks, and performance expectations as part of a job analysis of student work-study jobs on a college campus of 2,000 students, where 1,500 are involved in the work-study program.
- 6 You decide to take a telephone survey of 40 families in the 721-exchange area. You want an excellent representation of all subscribers in the exchange area. Explain how you will carry out this study.
- 7 You plan to conduct a mail survey of the traffic managers of 1,000 major manufacturing companies

across the country. The study concerns their company policies regarding the payment of moving expenses for employees who are transferred. What might you do to improve the response rate of such a survey?

- 8 A major corporation agrees to sponsor an internal study on sexual harassment in the workplace. This is in response to concerns expressed by its female employees. How would you handle the following issues:
- a The communication approach (self-administered, telephone, personal interview, and/or mixed).
  - b The purpose: Fact finding, awareness, relationship building, and/or change.
  - c Participant motivation.
  - d Minimization of response and nonresponse error.

**Bringing Research to Life**

- 9 Define the appropriate communication study for the Albany Outpatient Laser Clinic.

**From Concept to Practice**

- 10 Using Exhibit 10-1 as your guide, graph the communication study you designed in question 9.

>wwwexercise

Find a study that compares two methodologies, for example, telephone survey and Internet survey, or self-administered mail and Web survey. What does this study say about the effectiveness of the research methodologies being compared. One such study is at [http://www.mccic.org/reports\\_newsletters/year2001/FINAL\\_SPRING\\_Y01.pdf](http://www.mccic.org/reports_newsletters/year2001/FINAL_SPRING_Y01.pdf).

>cases\*

Can Research Rescue the Red Cross?

Donatos: Finding the New Pizza

Inquiring Minds Want to Know—NOW!

Mastering Teacher Leadership

NCRCC: Teeing Up and New Strategic Direction

Sturjel Division

USTA: Come Out Swinging

\* All cases appear on the text CD; you will find abstracts of these cases in the Case Abstracts section of this text. Video cases are indicated with a video icon.

# >chapter 11

## Experiments and Test Markets

“Your success in life isn’t based on your ability to simply change. It is based on your ability to change faster than your competition, customers and business.”

*Mark Sanborn, founder, Speakers Roundtable*

### >learning objectives

After reading this chapter, you should understand . . .

- 1 The uses for experimentation.
- 2 The advantages and disadvantages of the experimental method.
- 3 The seven steps of a well-planned experiment.
- 4 Internal and external validity with experimental research designs.
- 5 The three types of experimental designs and the variations of each.
- 6 The functions and types of test markets.

## >bringingresearchtolife

Jason slides into a seat next to Sally at the conference lunch table. Sally glances his way and nods briefly but keeps her attention on the man to her left. He is describing some of the finer details of the Point of Purchase Advertising Institute's groundbreaking experiment to put display materials on a directly comparative basis with other audited advertising and sales promotion activities. Sally had just attended his presentation during the morning session of the conference.

As he stops speaking, Sally introduces Jason, "Doug Adams, I'd like you to meet my colleague, Jason Henry. Jason, Doug is vice president and co-founder of Prime Consulting Group, Inc. [Prime]."

Jason extends his hand across Sally's plate, "Pleased to meet you, Doug. Prime took the lead on that POP industry experiment, right? Sorry I missed your presentation, but Sally" Jason nods to Sally, "suggested we divide and conquer—to cover more sessions."

Doug returns Jason's smile and handshake. "It seems to be the topic of conversation at this table at the moment, so maybe you'll get some of the content here," welcomes Doug. "I was just explaining how we needed a methodology that could separate out other sales influencers, like price, local advertising, media-delivered coupons, or a secondary stocking location in order to measure the sales lift generated by the POP material. For example, if Frito-Lay offered Doritos at \$2.49, reduced from \$2.99, plus it stocked a secondary location near the soft-drink aisle, Frito-Lay could track the sales lift.

If sales increased even more when a Doritos sign was posted over the secondary stock location, then the power of the POP could be determined."

"Your session was generating all the buzz as I entered the dining room," comments Jason. "How many types of POP were assessed?"

"Ultimately, 20 different types," shares Doug. "Several different message types—for example, brand name, photo, price, retail savings, thematic . . . like a movie tie-in . . . or generic . . . like the summer barbeque season—and numerous locations."

"Like regular shelf stocking location, end-cap, front lobby . . . ?" asks Jason. At Doug's affirmative nod Jason asks, "How did you keep the manufacturers from distorting the experiment?"

To give Doug a chance to take a bite, Sally supplies, "Prime used a double-blind audit tracking procedure using observation, coupled with sales tracking through more than 250 supermarkets from the IRI panel of stores and 120 convenience stores from six retailers with Nielsen's Market Decisions program. The manufacturers didn't know which stores were involved."

"And did the POP create the sales lift the retailers expected?" asks Jason.

"Not only were we able to calculate sales lift for each type of promotion," shares Doug, "but we were also able to calculate a full cost-per-thousand [CPM] estimate, including the cost of manufacturing, delivering and installing the point-of-purchase material. TV, radio, and in-store ads are still quoting CPM exposures without the cost of the ad."

“If grocery and convenience stores buy in, this will be a large piece of business for one or several firms,” comments Sally. “What’s been the reaction?”

“When a retailer, who has been relying on his gut instinct to accept or reject POP materials for his store, sees that that same material can cause a 20 to 40 percent lift in store sales, he’s bound to be receptive. We think it’s going to be a major new research initiative,” smiles Doug. “Are you interested?”

## > What Is Experimentation?

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

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

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

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

An example of such an intervention is the study of bystanders and thieves.<sup>1</sup> In this experiment, participants were asked to come to an office where they had an opportunity to see a person steal some money from a receptionist’s desk. A confederate of the experimenter, of course, did the stealing. The major hypothesis concerned whether people observing a theft will be more likely to report it (1) if they are alone when they observe the crime or (2) if they are in the company of someone else.

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

On what grounds did the researchers conclude that people who were alone were more likely to report crimes observed than people in the company of others? Three types of evidence form the basis for this conclusion. First, there must be an agreement between independent and dependent variables. The presence or absence of one is associated with the presence or absence of the other. Thus, more reports of the theft (DV) came from lone observers (IV<sub>1</sub>) than from paired observers (IV<sub>2</sub>).



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

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

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

## > An Evaluation of Experiments

### Advantages

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

The second advantage of the experiment is that contamination from extraneous variables can be controlled more effectively than in other designs. This helps the researcher isolate experimental variables and evaluate their impact over time. Third, the convenience and cost of experimentation are superior to other methods. These benefits allow the experimenter opportunistic scheduling of data collection and the flexibility to adjust variables and conditions that evoke extremes not observed under routine circumstances. In addition, the experimenter can assemble combinations of variables for testing rather than having to search for their fortuitous appearance in the study environment.

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

### Disadvantages

The artificiality of the laboratory is arguably the primary disadvantage of the experimental method. However, many subjects' perceptions of a contrived environment can be improved by investment in the facility. Second, generalization from nonprobability samples can pose problems despite random assignment. The extent to which a study can be generalized from college students to managers or executives is open to question. And when an experiment is unsuccessfully disguised, volunteer subjects are often those with the most interest in the topic. Third, despite the low costs of experimentation, many applications of experimentation far outrun the

budgets for other primary data collection methods. Fourth, experimentation is most effectively targeted at problems of the present or immediate future. Experimental studies of the past are not feasible, and studies about intentions or predictions are difficult. Finally, management research is often concerned with the study of people. There are limits to the types of manipulation and controls that are ethical.

## > Conducting an Experiment<sup>2</sup>

In a well-executed experiment, researchers must complete a series of activities to carry out their craft successfully. Although the experiment is the premier scientific methodology for establishing causation, the resourcefulness and creativeness of the researcher are needed to make the experiment live up to its potential. In this section, and as we introduce Exhibit 11-1, we discuss seven activities the researcher must accomplish to make the endeavor successful:

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

## Selecting Relevant Variables

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

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

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

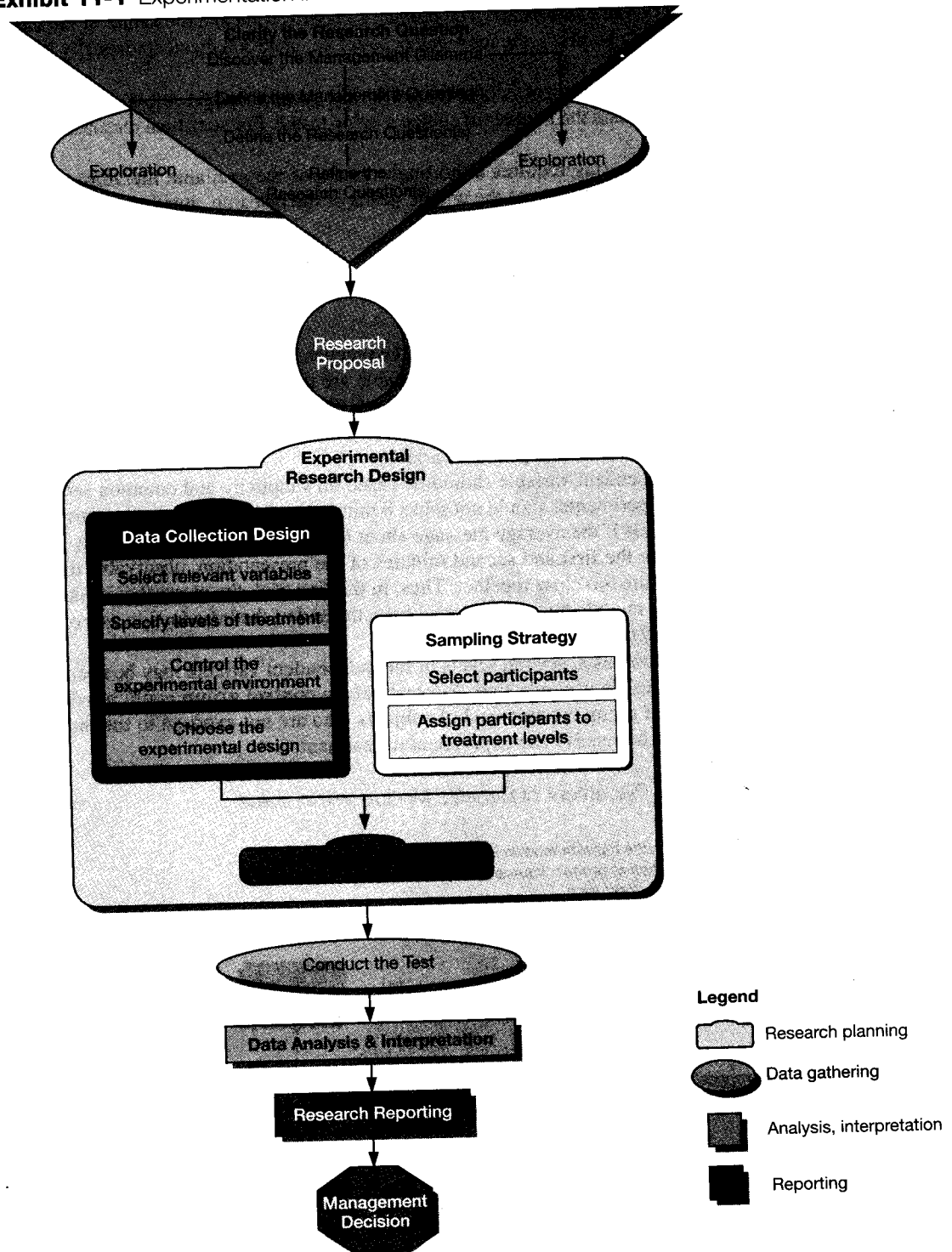
Since a hypothesis is a tentative statement—a speculation—about the outcome of the study, it might take this form:

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

The researchers' challenges at this step are to:

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

> Exhibit 11-1 Experimentation in the Research Process



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

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

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

## Specifying Treatment Levels

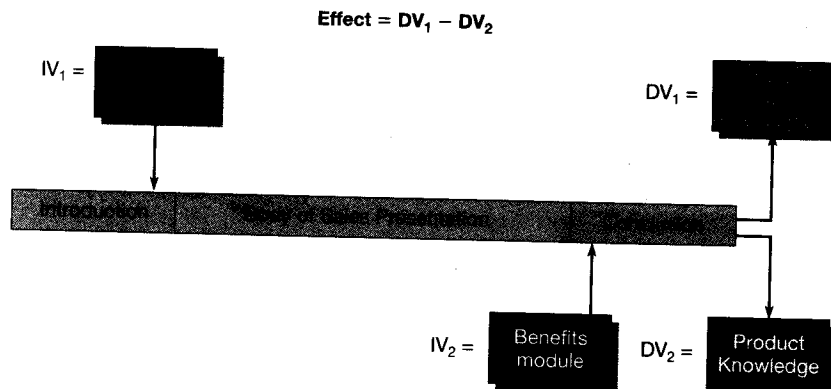
In an experiment, participants experience a manipulation of the independent variable, called the **experimental treatment**. The **treatment levels** of the independent variable are the arbitrary or natural groups the researcher makes within the independent variable of an experiment. For example, if salary is hypothesized to have an effect on employees' exercising of stock purchase options, it might be divided into high, middle, and low ranges to represent three levels of the independent variable.

The levels assigned to an independent variable should be based on simplicity and common sense. In the sales presentation example, the experimenter should not select 8 minutes and 10 minutes as the starting points to represent the two treatment levels if the average message about the product is 12 minutes long. Similarly, if the benefits module is placed in the first and second minutes of the presentation, observable differences may not occur because the levels are too close together. Thus, in the first trial, the researcher is likely to position the midpoint of the benefits module the same interval from the end of the introduction as from the end of the conclusion (see Exhibit 11-2).

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

### > Exhibit 11-2 Experiment of Placement of Benefits Module within Sales Presentation

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



## Controlling the Experimental Environment

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

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

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

## Choosing the Experimental Design

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

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

## Selecting and Assigning Participants

The participants selected for the experiment should be representative of the population to which the researcher wishes to generalize the study's results. This may seem self-evident, but we have witnessed several decades of experimentation with college sophomores that contradict that assumption. In the sales presentation example, corporate buyers, purchasing managers, or others in a decision-making capacity would provide better generalizing power than undergraduate college students *if* the product in question was targeted for industrial use rather than to the consumer.

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

When it is not possible to randomly assign subjects to groups, matching may be used. **Matching** employs a nonprobability quota sampling approach. The object of matching is to have each experimental and control

## &gt;snapshot

## Unicast's Video Ad Outperforms TV Ad

Advertising on the Internet has experienced mixed results, and as a result, most advertisers still rely heavily on television ads. But Unicast, an online advertising provider to more than 1,400 Web partners and 600 ad agencies and advertisers, has the research to prove its newest Internet advertising technology—a 14-million-mile-per-hour video download—can deliver on the most important ad placement metrics.

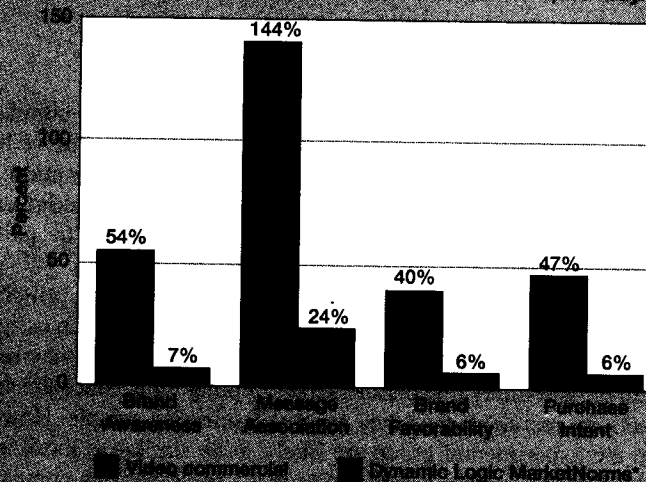
For a six-week period (January 18 through February 29, 2004) Unicast tested its proprietary video commercial format with ads from five advertisers, including AT&T, Honda, Pamp, Vonage, and World Group. The ads ran on 10 sites and were tested on a site including ABCnews.com, ESPN.com, Lycos, Adult.com, and MSN.com. During the test, 3,592 Internet users received a pop-up invitation to participate in a Web survey. 1,745 had experienced a one-line exposure to one of the video commercial ads, while 1,847 were not exposed to any of the ads. Dynamic Logic conducted the research and measured the results against its Admeter WebSite Data indicated an ad tested through both quarters 2002, n = 1,204, 1997. "We've been monitoring annoyance with online ads for three years," stressed Unicast marketing director Annette Muth. "So really, we were concerned that annoyance might increase with a 30-second ad."

The survey reports that while 28 percent found the video ads annoying, this is significantly less than the 35 percent who find television ads annoying, and overwhelmingly less than the 75.3 percent of consumers who find pop-up ads very annoying. More importantly, the video commercial format vastly outperformed conventional online ads (banners, pop-ups, etc.) on metrics like brand awareness, message association, brand favorability, and purchase intent.

Unicast's video commercial plays 30-second ads in broadcast quality (30 frames per second). It is precached (downloads fully before seen by the viewer) and interactive (can contain click-through options from the ad). "Vonage, the advertiser in the test that had the lowest awareness going in saw the most to gain," explained Muth. "experienced high percentage increases across all metrics." Advertisers were possibly surprised by the results. As a result more Web sites and advertisers have signed up to use the video format ads that 69 percent of test participants indicated were "like watching a TV ad" on their computer.

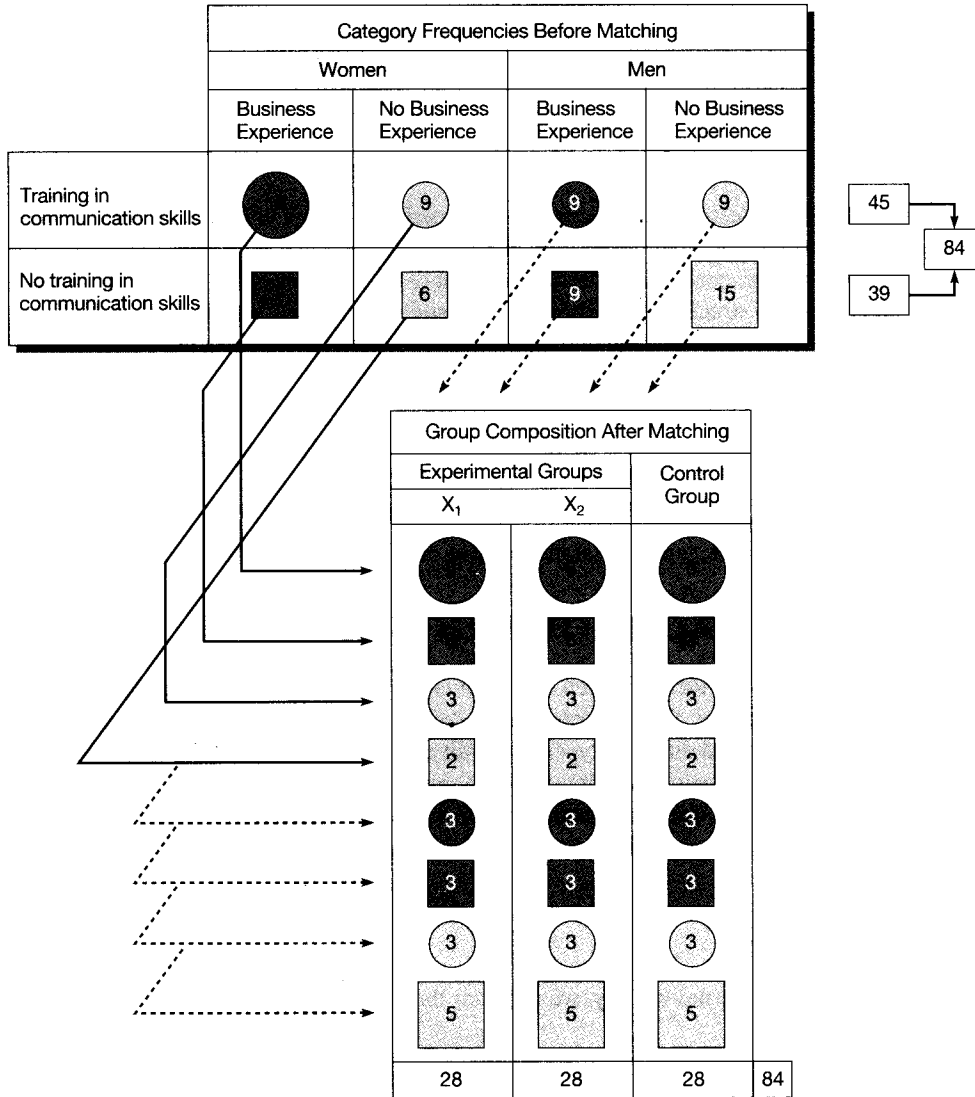
Visit Unicast's Web site gallery to see different types of Internet ads and the research that supports their use by advertisers.

[www.unicast.com](http://www.unicast.com); [www.dynamiclogic.com](http://www.dynamiclogic.com)



subject matched on every characteristic used in the research. This becomes more cumbersome as the number of variables and groups in the study increases. Since the characteristics of concern are only those that are correlated with the treatment condition or the dependent variable, they are easier to identify, control, and match.<sup>4</sup> In the sales presentation experiment, if a large part of the sample was composed of businesswomen who had recently completed communications training, we would not want the characteristics of gender, business experience, and communication training to be disproportionately assigned to one group.

> Exhibit 11-3 Quota Matrix Example



Some authorities suggest a **quota matrix** as the most efficient means of visualizing the matching process.<sup>5</sup> In Exhibit 11-3, one-third of the subjects from each cell of the matrix would be assigned to each of the three groups. If matching does not alleviate the assignment problem, a combination of matching, randomization, and increasing the sample size would be used.

## Pilot Testing, Revising, and Testing

The procedures for this stage are similar to those for other forms of primary data collection. Pilot testing is intended to reveal errors in the design and improper control of extraneous or environmental conditions. Pretesting the instruments permits refinement before the final test. This is the researcher's best opportunity to revise scripts, look for control problems with laboratory conditions, and scan the environment for factors that might confound

the results. In field experiments, researchers are sometimes caught off guard by events that have a dramatic effect on subjects: the test marketing of a competitor's product announced before an experiment, or a reduction in force, reorganization, or merger before a crucial organizational intervention. The experiment should be timed so that subjects are not sensitized to the independent variable by factors in the environment.

## Analyzing the Data

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

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

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

## > Validity in Experimentation

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

## Internal Validity

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

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

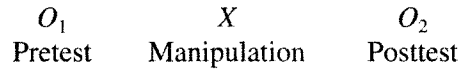
## History

During the time that an experiment is taking place, some events may occur that confuse the relationship being studied. In many experimental designs, we take a control measurement ( $O_1$ ) of the dependent variable



before introducing the manipulation ( $X$ ). After the manipulation, we take an after-measurement ( $O_2$ ) of the dependent variable. Then the difference between  $O_1$  and  $O_2$  is the change that the manipulation has caused.

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



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

### Maturation

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

### Testing

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

### Instrumentation

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

### Selection

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

### Statistical Regression

This factor operates especially when groups have been selected by their extreme scores. Suppose we measure the output of all workers in a department for a few days before an experiment and then conduct the experiment with only those workers whose productivity scores are in the top 25 percent and bottom 25 percent. No

matter what is done between  $O_1$  and  $O_2$ , there is a strong tendency for the average of the high scores at  $O_1$  to decline at  $O_2$  and for the low scores at  $O_1$  to increase. This tendency results from imperfect measurement that, in effect, records some persons abnormally high and abnormally low at  $O_1$ . In the second measurement, members of both groups score more closely to their long-run mean scores.

## Experiment Mortality

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

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

### ethics? ISSUES

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

## External Validity

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

- Reactivity of testing on  $X$ .
- Interaction of selection and  $X$ .
- Other reactive factors.

### The Reactivity of Testing on $X$

The reactive effect refers to sensitizing subjects via a pretest so that they respond to the experimental stimulus ( $X$ ) in a different way. A before-measurement of a subject's knowledge about the ecology programs of a

company will often sensitize the subject to various experimental communication efforts that might be made about the company. This before-measurement effect can be particularly significant in experiments where the IV is a change in attitude.

### Interaction of Selection and X

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

### Other Reactive Factors

The experimental settings themselves may have a biasing effect on a subject's response to X. An artificial setting can obviously produce results that are not representative of larger populations. Suppose the workers who



>picprofile

Researchers know that as many as 60 percent of purchase decisions are made in the store. Thus marketers aggressively seek in-store space to place temporary displays, shelf-talkers, and instant coupons, as well as ceiling signs and banners. Even the floor is contested real estate. So the ability to demonstrate the effectiveness of promotional materials is critical. FLOORgraphics, Inc., uses a longitudinal design, tracking sales of products in matched groups of stores (test and control groups). After test stores receive the FLOORad, relative sales in both groups are again compared to pre-ad performance and to each other. Research shows the FLOORad effect (the percentage sales increase directly due to the FLOORad) can lift sales 30 to 40 percent depending on the product category.

www.floorgraphics.com

are given the incentive pay are moved to a different work area to separate them from the control group. These new conditions alone could create a strong reactive condition.

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

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

## > Experimental Research Designs

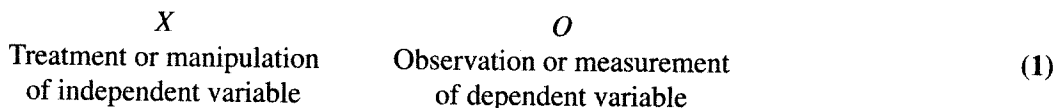
The many experimental designs vary widely in their power to control contamination of the relationship between independent and dependent variables. The most widely accepted designs are based on this characteristic of control: (1) preexperiments, (2) true experiments, and (3) field experiments (see Exhibit 11-4).

### Preexperimental Designs

All three preexperimental designs are weak in their scientific measurement power—that is, they fail to control adequately the various threats to internal validity. This is especially true of the after-only study.

#### After-Only Study

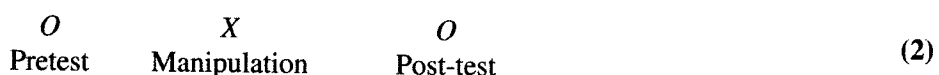
This may be diagrammed as follows:



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

#### One-Group Pretest-Posttest Design

This is the design used earlier in the educational example. It meets the various threats to internal validity better than the after-only study, but it is still a weak design. How well does it control for history? Maturation? Testing effect? The others?



> Exhibit 11-4 Key to Design Symbols

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

The Xs and Os in the diagram are read from left to right in temporal order.

O X O O

Time →

When multiple Xs and Os appear vertical to each other, this indicates that the stimuli and/or the observations take place simultaneously.

X O

O

Time →

q

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

X O

O

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

O X O

---

O

### Static Group Comparison

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

$$\begin{array}{ccc}
 X & & O_1 \\
 \text{-----} & & \text{-----} \\
 & & O_2
 \end{array} \tag{3}$$

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

## &gt;snapshot

Vanguard, a major provider of retirement benefit programs, is conducting an experiment within Philips Electronics North America to determine whether employees can be encouraged to increase the amount they save in their 401(k) retirement plans. When asked if they could increase their savings, most employees indicated that they live "paycheck to paycheck" and therefore cannot save more. Yet financial planners know that most people can save 1 percent, 3 percent, or even 5 percent more of their income over time and not notice a difference in their standard of living. The Vanguard/Philips experiment attempts to overcome this "painful to save" barrier by having workers agree to save more in the *future*—not today. In the experiment, which began in February 2002, about 600 workers in two geographically separate and distinct divisions of Philips (D<sub>1</sub> and D<sub>2</sub>) have been invited to join the SmartSave program. Under the program, they have the choice of increasing their 401(k) savings rate by 1, 2, or 3 percent drawn from a future pay increase. The rate change will occur on April 1 of each year, at the time of future merit increases. Whatever rate they choose, that increase will occur each April during the life of the experiment, unless they decide to discontinue or increase their savings rate.

SmartSave is being introduced with lots of fanfare, including a newsletter, two teaser postcards, SmartSave posters in

the workplace, a required attendance meeting on company time in which the program will be explained, and company raffles for participants. Additionally, workers in D<sub>2</sub> are being offered one-on-one meetings with a local financial planner. Vanguard and Philips will analyze several pre- and postmetrics:

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

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

[www.philips.com](http://www.philips.com); [www.vanguard.com](http://www.vanguard.com)

## True Experimental Designs

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

It is common to show an  $X$  for the test stimulus and a blank for the existence of a control situation. This is an oversimplification of what really occurs. More precisely, there is an  $X_1$  and an  $X_2$ , and sometimes more. The  $X_1$  identifies one specific independent variable, while  $X_2$  is another independent variable that has been chosen, often arbitrarily, as the control case. Different levels of the same independent variable may also be used, with one level serving as the control.

### Pretest-Posttest Control Group Design

This design consists of adding a control group to the one-group pretest-posttest design and assigning the subjects to either of the groups by a random procedure ( $R$ ). The diagram is:

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

The effect of the experimental variable is

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

# >snapshot

Ever wonder how consumer product companies test the effectiveness of their creations? At Hill Top Research, Inc., founded in 1947 and the largest consumer product testing firm in the world, researchers use a variety of devices—including the human nose. In one deodorant study subjects were brought to a test site that contained a *hot room*. Researchers applied the product being tested to each subject's armpit, followed by the insertion of a cotton pad under each arm, which subjects retained by pressing their arms to their sides. Researchers then led subjects to the *hot room*—where temperatures are warm enough to make anyone sweat. When the subjects exit the room after the defined

period of time, the cotton pad was removed for analysis. Then the odor detective did his or her job. A cup with a small hole in the bottom was placed against the subject's armpit (to assure uniform distance between nose and pit) and then the detective positioned her nose near the hole and inhaled. With a successful formulation, the odor detective would not detect a strong or offensive odor.

What are some of the variables a researcher would need to control in this study? What sources of error must be controlled?

[www.hill-top.com](http://www.hill-top.com)

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

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

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

## Posttest-Only Control Group Design

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

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

The experimental effect is measured by the difference between  $O_1$  and  $O_2$ :

$$E = (O_2 - O_1)$$

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

## Field Experiments: Quasi- or Semi-Experiments<sup>8</sup>

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

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

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

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

### Nonequivalent Control Group Design

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

$$\begin{array}{ccc}
 O_1 & X & O_2 \\
 \hline
 O_3 & & O_4
 \end{array} \tag{6}$$

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

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

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

### Separate Sample Pretest-Posttest Design

This design is most applicable when we cannot know when and to whom to introduce the treatment but we can decide when and whom to measure. The basic design is:



# >snapshot

With 70 percent of new product introductions failing, the test market, a long-standing experimental research tradition in the marketing research of consumer goods, is under attack. A test market involves placing a new product in a sample of stores, usually in two or more cities, and then monitoring sales performance under different promotion, pricing, and physical placement conditions. Nonproponents claim the racial diversity of the U.S. population is not well reflected in several U.S. cities often chosen to provide crucial evidence of product viability. Cedar Rapids (Iowa), Columbus (Ohio), Little Rock (Arkansas), and Evansville (Indiana), four popular test market locations, are primarily populated by non-Hispanic white households. Yet 22.9 percent of U.S. households claimed a different ethnic background in the Census Bureau's 2000 decennial census.

Information Resources, Inc. (IRI), whose *BehaviorScan* syndicated research uses similar cities, claims household demographics are only one criterion for site choice. Valerie Skala, IRI's vice president of Analytic Product Management and Development, claims "category and brand purchase patterns, as well as representative retail development," are more important. Test marketing site choice is also determined by cost and availability of retailers willing to stock the new product. IRI and VNU's ACNielsen Market Decisions have contracts with retailers that facilitate immediate placement, eliminating the need to negotiate placement with each store's corporate office. Does the absence of comparable demographics in test markets render the test market methodology faulty?

[www.vnu.com](http://www.vnu.com); [www.iri.com](http://www.iri.com)

$$\begin{array}{rcc}
 R & O_1 & (X) \\
 R & & X \quad O_2
 \end{array} \quad (7)$$

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

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

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

## Group Time Series Design

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

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

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

## > Test Marketing

This section examines traditional and emerging designs for test marketing including the characteristics of six test market types and the strengths and weaknesses of each type.

A **test market** is a controlled experiment conducted in a carefully chosen marketplace (e.g., Web site, store, town, or other geographic location) to measure marketplace response and predict sales or profitability of a product. The objective of a market test study is to assist marketing managers introduce new products or services, add products to existing lines, identify concepts with potential, or relaunch enhanced versions of established brands. By testing the viability of a product, managers reduce the risks of failure. Complex experimental designs are often required to meet the controlled experimental conditions of test markets. They also are used in other research where control of extraneous variables is essential. We describe the extensions of true experimental designs in this chapter's appendix.

The successful introduction of new products is critical to a firm's financial success. Failures not only create significant losses for companies but also hurt the brand and company reputation. According to ACNielsen, the failure rate for new products approaches 70 percent.<sup>9</sup> Estimates from other sources vary between 40 and 90 percent depending on whether the products are in consumer or industrial markets. Product failure may be attributable to many factors, especially inadequate research. Test-marketed products, typically evaluated in consumer industries, enjoy a significantly higher success rate because managers can reduce their decision risk through reality testing. They gauge the effectiveness of pricing, packaging, promotions, distribution channels, dealer response, advertising copy, media usage patterns, and other aspects of the marketing mix. Test markets also help managers evaluate improved versions of existing products and services.

### Test Market Selection

There are several criteria to consider when selecting test market locations. As we mentioned earlier, one of the primary advantages of a carefully conducted experiment is external validity or the ability to generalize to (and across) times, settings, or persons. The location and characteristics of participants should be *representative* of the market in which the product will compete. This requires consideration of the product's target competitive environment, market size, patterns of media coverage, distribution channels, product usage, population size, housing, income, lifestyle attributes, age, and ethnic characteristics. Not even "typical" all-American cities are ideal for all market tests. Kimberly-Clark's Depend and Poise brand products for bladder control could not be adequately tested in a college town. Cities that are overtested create problems for market selection because savvy participants' prior experiences cause them to respond atypically.

*Multiple locations* are often required for optimal demographic balance. Sales may vary by region, necessitating test sites that have characteristics equivalent to those of the targeted national market. Several locations may also be required for experimental and control groups.

*Media coverage* and *isolation* are additional criteria for locating the test. Although the test location may not be able to duplicate precisely a national media plan, it should adequately represent the planned promotion through print and broadcast coverage. Large metropolitan areas produce media spillover that may contaminate the test area. Advertising is wasted as the media alerts distributors, retailers, and consumers in adjacent areas about the product. Competitors are warned more quickly about testing activities and the test loses its competitive advantage. In 2002, Dairy Queen (DQ) Corp., which has 5,700 stores throughout the world, began testing electronic irradiated burgers at the Hutchinson and Spicer locations in Minnesota. No quick-service restaurant chains provide irradiated burgers, although McDonald's and Burger King also researched this

option. DQ originally focused information about the test at the store level rather than with local media. When the *Minneapolis Star Tribune* ran a story about the test, DQ had to inform all Minnesota store operators about the article, although all operators had known about the planned test. The article created awareness for anti-irradiation activists and the potential for demonstrations—an unplanned consequence of the test market.<sup>10</sup> Although relatively isolated communities are more desirable because their remoteness aids controlling critical promotional features of the test, in this instance media spillover and unintended consequences of unplanned media coverage became a concern.

The *control of distribution* affects test locations and types of test markets. Cooperation from distributors is essential for market tests conducted by the product's manufacturer. The distributor should sell exclusively in the test market to avoid difficulties arising from out-of-market warehousing, shipping, and inventory control. When distributors in the city are either unavailable or uncooperative, a controlled test, where the research firm manages distribution, should be considered.

## Types of Test Markets

There are six major types of test markets: standard, controlled, electronic, simulated, virtual, and Web-enabled. In this section, we discuss their characteristics, advantages and disadvantages, and future uses.

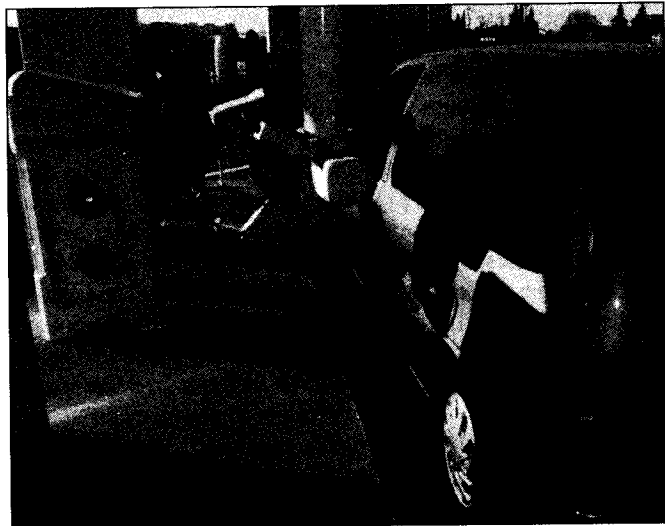
### Standard Test Market

The **standard test market** is a traditional test of a product and/or marketing mix variables on a limited geographic basis. It provides a real-world test for evaluating products and marketing programs on a smaller, less costly scale. The firm launching the product selects specific sales zones, test market cities, or regions that have characteristics comparable to those of the intended consumers of the product. The firm performs the test through its existing distribution channels, using the same elements as used in a national rollout. Exhibit 11-5 shows some U.S. cities commonly used as test markets.

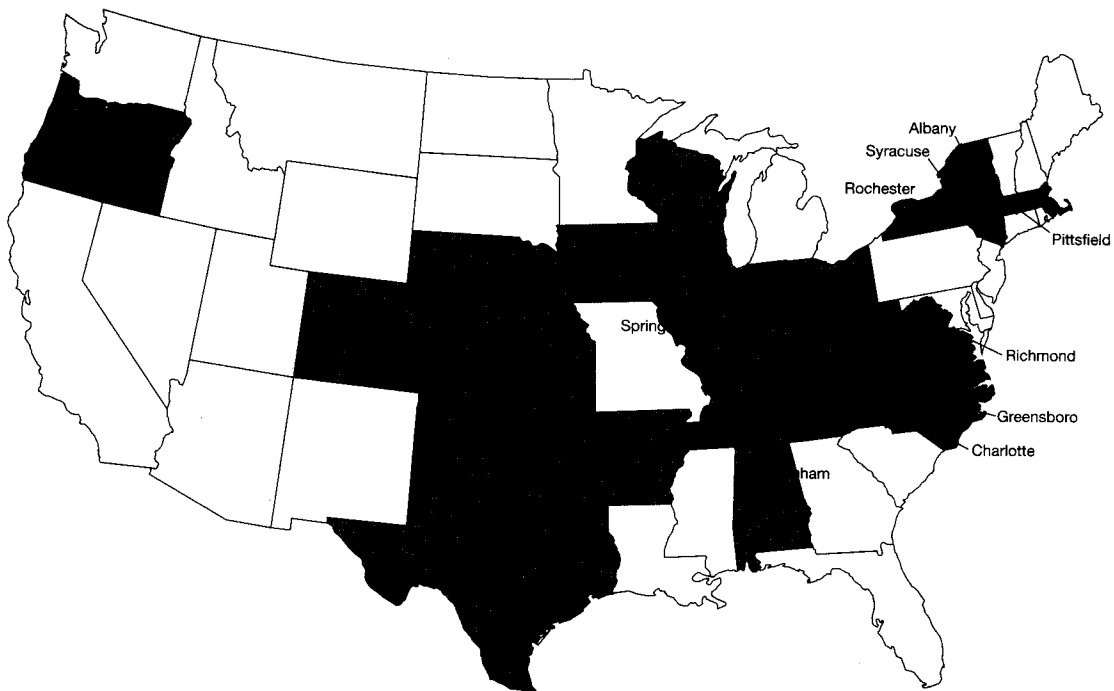
Standard test markets benefit from using actual distribution channels and discovering the amount of trade support necessary to launch and sustain the product. High costs (\$1 million is typical, ranging upward to \$30 million) and long time (12 to 18 months for a go/no-go decision) are disadvantages. The loss of secrecy when the test exposes the concept to the competition further complicates the usefulness of traditional tests.

In March 2000, in an affluent suburb of Indianapolis, Shell Oil Co. test-marketed the first robotic gas pump that allows drivers to serve themselves without leaving their cars. The innovation, which uses a combination of robotics, sensors, and cameras to guide the fuel nozzle into a vehicle's gas tank, took eight years to develop. Its features allow a parent to stay with children while pumping gas and enable a driver to avoid exposure to gas fumes or the risk of spillage, static fire, or even bad weather. Unfortunately, the product requires a coded computer chip containing vehicle information that must be placed on the windshield and a special, spring-loaded gas cap, which costs \$20. The introduction could hardly

The SmartPump is a robotic gas pump that dispenses fuel without the customer ever getting out of the car. Customers pay an additional \$1 for the service. [www.shell.com](http://www.shell.com)



## &gt; Exhibit 11-5 Test Market Cities



Source: Acxiom Corporation, a database services company, released its first "Mirror on America" May 24, 2004, ranking America's top 150 Metropolitan Statistical Areas (MSAs) on overall consumer test market characteristics. "Which American City Provides the Best Consumer Test Market?" [http://www.acxiom.com/default.aspx?ID=2521&Country\\_Code=USA](http://www.acxiom.com/default.aspx?ID=2521&Country_Code=USA). Also see <http://www.bizjournals.com/phoenix/stories/2000/11/20/daily5.html> and <http://celebrity-network.net/trc/business.htm>.

have been more ill-timed. Just as gasoline prices began their upward advance and the end of winter removed the incentive for staying behind the wheel, Shell planned to charge an extra \$1 per fill-up.<sup>11</sup>

## Controlled Test Markets

The term **controlled test market** refers to real-time forced distribution tests conducted by a specialty research supplier that guarantees distribution of the test product through outlets in selected cities. The test locations represent a proportion of the marketer's total store sales volume. The research firm typically handles the retailer sell-in process and all distribution activities for the client during the market test. The firm offers financial incentives for distributors to obtain shelf space from nationally prominent retailers and provides merchandising, inventory, pricing, and stocking control. Using scanner-based, survey, and other data sources, the research service gathers sales, market share, and consumer demographics data, as well as information on first-year volumes.

Companies such as ACNielsen Market Decisions and Information Resources, Inc., give consumer packaged-goods (CPG) manufacturers the ability to evaluate sales potential while reducing the risks of new or relaunched products prior to a national rollout. Market Decisions, for example, has over 25 small to medium-size test markets available nationwide. Typically, consumers experience all the elements associated with the first-year marketing plan, including media advertising and consumer and trade promotions. Manufacturers with a substantial commitment to a national rollout also have the opportunity to "fast-track" products during a condensed time period (three to six months) before launch.<sup>12</sup>

*Consumer packaged goods are consumer goods packaged by manufacturers and not sold unpackaged (in bulk) at the retail level (e.g., food, drinks, personal care products).*

Controlled test markets cost less than traditional ones (although they may reach several million dollars per year). They reduce the likelihood of competitor monitoring and provide a streamlined distribution function through the sponsoring research firm. Their drawbacks include the number of markets evaluated, the use of incentives—which distort trade cost estimates—and the evaluation of advertising.

## Electronic Test Markets

An **electronic test market** is a test system that combines store distribution services, consumer scanner panels, and household-level media delivery in specifically designated markets. Retailers and cable TV operators have cooperative arrangements with the research firm in these markets. Electronic test markets, previously used with consumer packaged-goods brands, have the capability to measure marketing mix variables that drive trial and repeat purchases by demographic segment for both CPG and non-CPG brands. Information Resources Inc. (IRI), for example, offers a service called BehaviorScan, which is also known as a *split-cable test* or *single-source test*, that combines scanner-based consumer panels with sophisticated broadcasting systems. IRI uses a combination of Designated Market Area-level cut-ins on broadcast networks and local cable cut-ins to assess the effect of the advertising that the household panel views. IRI and ACNielsen collect supermarket, drugstore, and mass merchandiser scanner data used in such systems. The BehaviorScan service makes use of these data with respondents who are then exposed to different commercials with various advertising weights.<sup>13</sup>

IRI's TV system operates as a within-market TV advertising testing service. The five BehaviorScan markets are Eau Claire, Wisconsin; Cedar Rapids, Iowa; Midland, Texas; Pittsfield, Massachusetts; and Grand Junction, Colorado. As small markets, with populations of 75,000 to 215,000, they provide lower marketing support costs than other test markets and offer appropriate experimental controls over the test conditions. Although several thousand households may be used, by assigning every local cable subscriber a cell, the service can indiscernibly deliver different TV commercials to each cell and evaluate the effect of the advertising on the panelists' purchasing behavior. For a control, nonpanelist households in the cable cell are interviewed by telephone.

BehaviorScan tracks the actual purchases of a household panel through bar-coded products at the point of purchase. Participants show their identification card at a participating store and are also asked to "report purchases from non-participating retailers, including mass merchandisers and supercenters, by using a handheld scanner at home."<sup>14</sup> Computer programs link the household's purchases with television viewing data to get a refined estimate ( $\pm 10$  percent) of the product's national sales potential in the first year. Consider the observation of a Frito-Lay senior vice president:

BehaviorScan is a critical component of Frito-Lay's go-to-market strategy for a couple of reasons. First, it gives us absolutely the most accurate read on the sales potential of a new product, and a well-rounded view of consumer response to all elements of the marketing mix. Second, BehaviorScan TV ad testing enables us to significantly increase our return on our advertising investment.<sup>15</sup>

The advantages of electronic test markets are apparent from the quality of strategic information provided but suffer from an artifact of their identification card data collection strategy: participants may not be representative.

## Simulated Test Markets

A **simulated test market (STM)** occurs in a laboratory research setting designed to simulate a traditional shopping environment using a sample of the product's consumers. STMs do not occur in the marketplace but are often considered a pretest before a full-scale market test. STMs are designed to determine consumer response to product initiatives in a compressed time period. A computer model, containing assumptions of how the new product would sell, is augmented with data provided by the participants in the simulation.

STMs have common characteristics: (1) Consumers are interviewed to ensure that they meet product usage and demographic criteria; (2) they visit a research facility where they are exposed to the test product and may be shown commercials or print advertisements for target and competitive products; (3) they shop in a simulated store environment (often resembling a supermarket aisle); (4) those not purchasing the product are offered free samples; (5) follow-up information is collected to assess product reactions and to estimate repurchase intentions; and (6) researchers combine the completed computer model with consumer reactions in order to forecast the likely trial purchase rates, sales volume, and adoption behavior prior to market entry.

When in-store variations are used, research suppliers select three to five cities representing the market where the product will be launched. They choose a mall with a high frequency of targeted consumers. In the mall, a simulated store in a vacant facility is stocked with products from the test category. Intercept interviews qualify participants for a 15-minute test during which participants view an assortment of print or television advertisements and are asked to recall salient features. Measures of new product awareness are obtained. With “dollars” provided by the research firm, participants may purchase the test product or any of the competing products. Advertising awareness, packaging, and adoption are assessed with a computer model, as in the laboratory setting. Purchasers may be offered additional opportunities to buy the product at a reduced price in the future.

STMs were widely adopted in the 1970s by global manufacturers as an alternative to standard test markets, which were considered more expensive, slower, and less protected. Although STM models continue to work somewhat well in today’s mass-market world, their effectiveness will diminish in the next decade as the one-to-one marketing environment becomes more diverse. To obtain forecast accuracy at the individual level, not just trial or repeat probabilities, STMs require individualized marketing plans to estimate different promotional and advertising factors for each person.<sup>16</sup>

M/A/R/C Research, Inc., has what it calls its *Assessor* model with many features that address the deficiencies of previous STM forecasting models. For example, instead of a comparison of consumer reactions to historical databases, individual consumer preferences and current experiences with existing brands help to define the fit for the new product environment. A competitive context pertinent to each consumer’s unique set of alternatives plays a prominent role in new product assessment. Important user segments (e.g., parent brand users, heavy users, or teenagers) are analyzed separately to capture distinct behaviors. According to M/A/R/C, the results of three different models (attitudinal preference models; a trial, repeat, depth-of-repeat model; and a behavioral decision model) are merged to reduce the influence of bias. From an accuracy standpoint, over 90 percent of the validated Assessor forecasts are within 10 percent of the actual, in-market sales volume figures.<sup>17</sup> Realistically, plus or minus 10 percent represents a level of precision that many firms are not willing to accept.

STMs offer several benefits. The cost (\$50,000 to \$150,000) is one-tenth of the cost of a traditional test market, competitor exposure is minimized, time is reduced to six to eight months, and modeling allows the evaluation of many marketing mix variables. The inability to measure trade acceptance and its lack of broad-based consumer response are its drawbacks.

## Virtual Test Markets

A **virtual test market** uses a computer simulation and hardware to replicate the immersion of an interactive shopping experience in a three-dimensional environment. Essential to the immersion experience is the system’s ability to render realistically product offerings in real time. Other features of interactive systems are the ability to explore (navigate in the virtual world) and manipulate the content in real time. In virtual test markets, the participants move through a store and display area containing the product. They handle the product by touching its image and examine it dimensionally with a rotation device to inspect labels, prices, usage instructions, and packaging. Purchases are made by placing the product in a shopping cart. Data collected

include time spent by product category, frequency and time with product manipulation, and order quantity and sequence, as well as video feedback of participant behavior.

An example of a virtual environment application reveals it as an inexpensive research tool:

Goodyear conducted a study of nearly 1,000 people. . . . Each respondent took a trip through a number of different virtual tire stores stocked with a variety of brands and models. . . . Goodyear found the results of the test valuable on several fronts. First, the research revealed the extent to which shoppers in different market segments valued the Goodyear brand over competing brands. Second, the test suggested strategies for repricing the product line.<sup>18</sup>

Virtual test markets are part of a family of virtual technology techniques dating back to the early 1990s. The term *Virtual Shopping*® was registered by Allison Research Technologies (ART) in the mid-90s.<sup>19</sup> ART's interfaces create a detailed virtual environment (supermarket, bar/tavern, convenience store, fast-food restaurant, drugstore, computer store, car dealership, and so forth) for participant interaction. Consumers use a display interface to point out what products are appealing or what they might purchase. Products, in CPG and non-CPG categories, are arrayed just as in an actual store. Data analysis includes the current range of sophisticated research techniques and simulated test market methodologies.<sup>20</sup> Improvements in virtual reality technology are creating opportunities for multisensory shopping. Current visual and auditory environments are being augmented with additional modes of sensory perception such as touch, taste, and smell.

A hybrid market test that bridges virtual environments and Internet platforms begins to solve the difficult challenge of product design teams: concept selection. A traditional reliance on expensive physical prototypes may be resolved with virtual prototypes. Virtual prototypes were discovered to provide results comparable to those of physical ones, cost less to construct, and allow Web researchers to explore more concepts. In some cases, however, the computer renderings make virtual prototypes look better in virtual reality and score lower in physical reality—specially when comparisons are made with commercially available products.<sup>21</sup>

## Web-Enabled Test Markets

Manufacturers have found an efficient way to test new products, refine old ones, survey customer attitudes, and build relationships. **Web-enabled test markets** are product tests using online distribution. They are primarily used by large CPG manufacturers that seek fast, cost-effective means for estimating new product demand. Although they offer less control than traditional experimental design, Procter & Gamble test-marketed Dryel, the home dry-cleaning product, for more than three years on 150,000 households in a traditional fashion while Drugstore.com tested the online market before its launch in 1999, taking less than a week and surveying about 100 people. Procter & Gamble now conducts 40 percent of its 6,000 product tests online. The company's annual research budget is about \$140 million, but it believes that figure can be halved by shifting research projects to the Internet.<sup>22</sup>

In 2000, when P&G geared up to launch Crest Whitestrips, a home tooth-bleaching kit, its high retail price created uncertainty. After an eight-month campaign offering the strips solely through the product's dedicated Web site, it sold 144,000 whitening kits online. Promoting the online sale, P&G ran TV spots, placed advertisements in lifestyle magazines, and sent e-mails to customers who signed up to receive product updates (12 percent of whom subsequently made a purchase). Retailers were convinced to stock the product, even at the high price. By timing the introduction with additional print and TV ad campaigns, P&G sold nearly \$50 million worth of Crest Whitestrips kits three months later.<sup>23</sup> P&G's success has been emulated by its competitors and represents a growing trend. General Mills, Quaker, and a number of popular start-ups have followed, launching online test-marketing projects of their own.

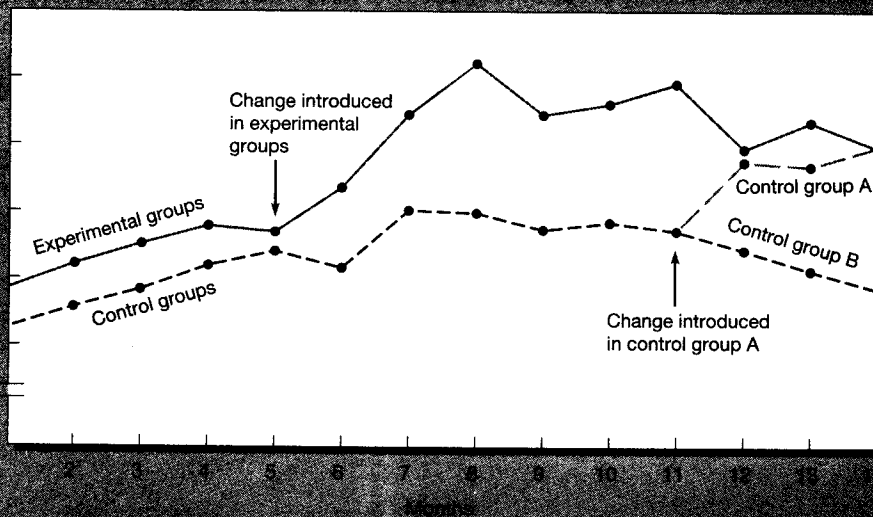
One theory of job satisfaction, Herzberg's "Two-Factor Theory," states that "hygiene factors" (such as salary, benefits, and working conditions) are necessary but not sufficient to create job satisfaction. "Motivators" (such as achievement, responsibility, and growth) are necessary for job satisfaction. This theory says that the positive motivator factors are intrinsic to the job; they include achievement, responsibility for achievement, the work itself, responsibility, and opportunity for advancement.

A study of the effects of job enrichment as a result of job redesign was carried out with laboratory technicians of Equipment Division (ED) at Ethicon Chemical. The object was to determine the effect of job enrichment on the job satisfaction of laboratory technicians. The study was conducted in two phases. In the first phase, the effect of job enrichment on job satisfaction was determined. In the second phase, the effect of job enrichment on job performance was determined. The study was conducted in two phases. In the first phase, the effect of job enrichment on job satisfaction was determined. In the second phase, the effect of job enrichment on job performance was determined. The study was conducted in two phases. In the first phase, the effect of job enrichment on job satisfaction was determined. In the second phase, the effect of job enrichment on job performance was determined.

The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period. The study continued over a 12-month period.

Several months of ED program development were completed prior to baseline for evaluation. The results of the study are shown in Exhibit 11-6.

EXHIBIT 11-6 Assessment of ED's Monthly Reports





## >summary

- 1 Experiments are studies involving intervention by the researcher beyond that required for measurement. The usual intervention is to manipulate a variable (the independent variable) and observe how it affects the subjects being studied (the dependent variable).  
An evaluation of the experimental method reveals several advantages: (1) the ability to uncover causal relationships, (2) provisions for controlling extraneous and environmental variables, (3) convenience and low cost of creating test situations rather than searching for their appearance in business situations, (4) the ability to replicate findings and thus rule out idiosyncratic or isolated results, and (5) the ability to exploit naturally occurring events.
- 2 Some advantages of other methods that are liabilities for the experiment include (1) the artificial setting of the laboratory, (2) generalizability from nonprobability samples, (3) disproportionate costs in select business situations, (4) a focus restricted to the present and immediate future, and (5) ethical issues related to the manipulation and control of human subjects.
- 3 Consideration of the following activities is essential for the execution of a well-planned experiment:
  - a Select relevant variables for testing.
  - b Specify the treatment levels.
  - c Control the environmental and extraneous factors.
  - d Choose an experimental design suited to the hypothesis.
  - e Select and assign subjects to groups.
  - f Pilot test, revise, and conduct the final test.
  - g Analyze the data.
- 4 We judge various types of experimental research designs by how well they meet the tests of internal and external validity. An experiment has high internal validity if one has confidence that the experimental treatment has been the source of change in the dependent variable. More specifically, a design's internal validity is judged by how well it meets seven threats. These are history, maturation, testing, instrumentation, selection, statistical regression, and experiment mortality.  
External validity is high when the results of an experiment are judged to apply to some larger population. Such an experiment is said to have high external validity regarding that population. Three potential threats to external validity are testing reactivity, selection interaction, and other reactive factors.
- 5 Experimental research designs include (1) preexperiments, (2) true experiments, and (3) quasi-experiments. The main distinction among these types is the degree of control that the researcher can exercise over validity problems.  
Three preexperimental designs were presented in the chapter. These designs represent the crudest form of

experimentation and are undertaken only when nothing stronger is possible. Their weakness is the lack of an equivalent comparison group; as a result, they fail to meet many internal validity criteria. They are the (1) after-only study, (2) one-group pretest-posttest design, and (3) static group comparison.

Two forms of the true experiment were also presented. Their central characteristic is that they provide a means by which we can ensure equivalence between experimental and control groups through random assignment to the groups. These designs are (1) pretest-posttest control group and (2) posttest-only control group.

The classical two-group experiment can be extended to multigroup designs in which different levels of the test variable are used as controls rather than the classical nontest control.

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

- 6 Test marketing is a controlled experimental procedure conducted in a carefully selected marketplace to test a product or service to predict sales and profit outcomes. Managers use test marketing to introduce new products or services, add products to existing lines, identify concepts with potential, or relaunch enhanced versions of established brands. There are six major types of test markets. A standard test market is a traditional test of a product and/or marketing mix variables on a limited geographic basis. It provides a real-world test on a smaller, less costly scale. The firm selects test market cities or regions comparable to those of the intended consumers of the product and tests it through its existing distribution channels. Controlled test markets are "live" forced distribution tests conducted by a specialty research supplier that guarantees distribution of the test product through outlets in selected cities. An electronic test market is a test system that combines store distribution services, consumer scanner panels, and household-level media delivery in specifically designated markets. Retailers and cable TV operators have cooperative arrangements with the research firm in these tests. A simulated test market (STM), often a pretest before a full-scale market test, occurs in a laboratory setting designed to simulate a traditional shopping environment

using a sample of the product's consumers. STMs use computer models and data provided by participants in the simulation. A virtual test market uses a computer simulation and hardware to replicate the immersion of an interactive shopping experience in a virtual, three-

dimensional environment. Web-enabled test markets are a growing trend for large consumer packaged-goods manufacturers that seek fast, cost-effective means to test new products, refine old ones, survey customer attitudes, and build relationships.

## >keyterms

<b>blind</b> 279	<b>hypothesis</b> 276	<b>test market</b> 292
<b>control group</b> 278	<b>independent variable (IV)</b> 274	<b>controlled test market</b> 294
<b>dependent variable (DV)</b> 274	<b>internal validity</b> 282	<b>electronic test market</b> 295
<b>double blind</b> 279	<b>matching</b> 279	<b>simulated test market (STM)</b> 295
<b>environmental control</b> 279	<b>operationalized</b> 276	<b>standard test market</b> 293
<b>experiment</b> 274	<b>quota matrix</b> 281	<b>virtual test market</b> 296
<b>experimental treatment</b> 278	<b>random assignment</b> 279	<b>Web-enabled test market</b> 297
<b>external validity</b> 282	<b>replication</b> 275	<b>treatment levels</b> 278
<b>field experiment</b> 275		

## >discussionquestions

### Terms in Review

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

### Making Research Decisions

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

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