Simulated Test Marketing

Technology for Launching Successful New Products

Kevin J. Clancy
Robert S. Shulman
Marianne M. Wolf

LEXINGTON BOOKS
An Imprint of Macmillan, Inc.
NEW YORK
Maxwell Macmillan Canada
TORONTO
Maxwell Macmillan International
NEW YORK OXFORD SINGAPORE SYDNEY
tive assumptions and estimated that the total marketing dollars spent by manufacturers for failed new food products in a single year ranged from $9 billion to $14 billion.17

"Even though the $9 billion to $14 billion dollar amount is huge in an absolute sense," said Malhotra, "by itself, it is hard to judge this number. To provide a perspective, I looked up the profits of the leading companies in the food business. The total food operating profit of the top fifteen companies was $7 billion in 1989. Thus, if my assumptions are correct, the 1989 marketing spending behind failed new food items may have been twice as much as the total profits of the top fifteen companies." Even the most conservative estimate indicates that companies spent more to introduce failed products than they earned—at least $2 billion more.

Malhotra saw three probable implications from his calculations.

1. The pace of new product introductions might slow. "If wholesalers and retailers are successful in tacking on a failure fee, that could certainly curb introductions of some close-in line extensions." A failure—or exit—fee increases the marketer's cost of doing business and ultimately consumer prices.

2. Retailers increasingly recognize that new products require more testing. Malhotra quoted Michael J. Rourke of A&P as saying, "When you have a 90 percent failure rate, you know there isn't enough testing going on." In the future, Malhotra speculated, retailers will probably demand more extensive results on new products before they accept and support them.

3. Marketers themselves increasingly recognize they must do more and better new product testing before they dump items onto the market to see whether they sell. Arthur Shulze, vice chairman of General Mills, has said, "The time is rapidly passing when a manufacturer can willy-nilly introduce a product without adequate testing."18

To reduce the huge cost and wasted effort on products that are likely to fail, companies have tried to reduce the risk by testing their new ideas before rolling them out nationally. As we'll see in a moment, however, traditional test marketing—the way most firms go about evaluating new products prior to a national introduction—may be more hindrance than help.

2

Traditional Test Marketing

A Recipe for Failure

Until fairly recently, companies that tested new products did so by conducting a real-world experiment. Indeed, David A. Aaker and George S. Day, the authors of one popular marketing research textbook, tell us that, "In terms of providing a realistic evaluation of a marketing program, there is no substitute for conducting an experiment in which the marketing program (or perhaps several versions of it) are implemented in a limited but carefully selected part of the market."1

The underlying assumption is that through such an experiment a company can determine a marketing program's impact with all its interdependencies in the actual market context. This is more valid and reliable than the artificial context of concept and product tests.

A company employs a test market for two basic reasons:

1. To gain information and experience with the marketing program before committing to it totally.
2. To predict the outcome when the company implements the marketing program in the national marketplace.

Gaining Information and Experience

The test market acts as a pilot operation for the national (or regional) marketing program, just as a pilot production facility might serve to debug a manufacturing process. As we saw with the Citrus
Hill Orange Juice example, marketing programs can develop a variety of problems. Here is another example of what can go wrong.

In the late 1980s, Weyerhaeuser Co., which has manufactured private label disposable diapers since 1970, decided that it had an opportunity to take a significant share of the $3.8 billion disposable diaper market. Weyerhaeuser would market its premium UltraSofts, which used a new technology to produce a softer, more absorbent diaper, as a hybrid combination of store label and national brand. In the Rochester, New York, test market where Wegmans Food Markets sold the line, for example, the package carried the name “Wegmans UltraSofts.” With this strategy, Weyerhaeuser could advertise the diapers over a large territory, like a national brand, but could use packaging and promotions tailored for each retailer. Weyerhaeuser also planned to sell UltraSofts for about 10 percent less than Procter & Gamble’s Pampers and Kimberly-Clark’s Huggies, which together command 85 percent of the disposable diaper market.

Weyerhaeuser hired marketing researchers to conduct a consumer test and the results showed that parents preferred the new UltraSofts two to one over the leading brands. With these encouraging results, Wegmans introduced UltraSofts with a huge promotion and advertising campaign, mailing 50,000 coupons to shoppers in the Rochester area. The effort was “unprecedented,” said Mary Ellen Burris, Wegmans’ consumer affairs director. With the $1-off coupon, a bag of 32 large diapers sold for $8.39, or about $1.60 less than Pampers and Huggies.

Unfortunately, when Weyerhaeuser began to manufacture the new diapers in volume, it discovered production line snags. “We’ve got about twenty problems,” reported Bobby Abraham, president of Weyerhaeuser’s personal care products division. “Many are simple. Some need significant work.” Because of the manufacturing problems, the factory could produce only enough product to supply Wegmans, and because it miscalculated the costs, Weyerhaeuser had to increase UltraSofts’ price to retailers 22 percent, wiping out its price advantage. Wegmans did maintain the $9.39 retail price, but the retailer could not maintain the advertising and promotional support, and its profit margins declined. In addition to Weyerhaeuser’s manufacturing headaches, it could never persuade other retailers in the Rochester area to stock UltraSofts because the price increase had made the brand less attractive.

Meanwhile, P&G and Kimberly-Clark spent heavily on the national rollout of their pink and blue diapers for girls and boys. Part of these efforts were thousands of $1- and $2-off coupons the companies sent to consumers.

This promotional effort made UltraSofts considerably less attractive to consumers. While neither Wegmans nor Weyerhaeuser will discuss exact numbers, Weyerhaeuser did say that ten months after introduction, UltraSofts’ market share was down considerably from its peak. “It’s been a disappointment,” said Abraham.

Real-World Tests Have Methodological Problems

If the test market’s first function is to gain information and experience with the marketing program before making a total commitment, the second is to predict the outcome when the company applies the marketing program to the total market. Although a real world test may provide a realistic measure of the marketing program’s impact, as we will see, it also has methodological problems that make market prediction difficult—and sometimes impossible.

A traditional test market has at least four steps:

1. The company selects the test cities.
2. It establishes the length of the test effort.
3. It implements and controls the marketing program in each city.
4. It undertakes research to evaluate the program and to forecast how the new brand would perform nationally.

Aaker and Day point out that the company must choose cities that meet several relevant characteristics. Ideally, a test market city should represent the country as a whole in terms of characteristics that will affect the test outcome; these include product usage, consumer attitudes, and demographics. If the researchers know the idiosyncrasies of a market, of course, they can adjust the results to compensate. For example, Southerners eat more biscuits than Americans on average; teenagers drink more soda; older people buy more pharmaceutical products.

The company must have access to data. Store audit information to evaluate a test is helpful because it provides sales data adjusted for inventory changes and gives other useful information, such as shelf facings and in-store promotions. If the company must have
store audit data, it must use cities that have retailers who will cooperate with store audits.

The test market communities should be isolated from contaminating media that "spills in" from nearby cities. For example, half the advertising reaching the Springfield-Holyoke, Massachusetts area originates elsewhere. On the other hand, too much media spilling out of the test market into surrounding areas is wasteful and increases the test's cost.

It may be desirable to use test market cities that do not have much "product spillage" outside the area so that it is possible to measure sales within the market. Columbus, Ohio, for example, has considerable spillage; Phoenix-Tucson has little.

After choosing the test market cities, the second major consideration is time. If possible, a test market normally should operate for one to two years. Even if after that year or two the company takes the program national, it should continue to monitor the test market to detect the impact of changes in the environment.

A market test requires an extended time for several reasons. Researchers can observe important seasonal factors only if the test continues for a full year. In addition, initial consumer interest often cannot predict a program's staying power. Finally, it is useful to allow the competition and other market factors to react to see what likely impact they will have on the new brand's performance.

The third major consideration is how the company carries out and controls the marketing program in each test city. Obviously the company wants the test cities to reflect the national program, but as we have already noted, this is never easy. The company may not have defined the national program precisely enough, or it may not be easy to apply the marketing program in selected test markets. For example, the company may not be able to translate the national advertising budget to a local level, or the test market may not be well defined in terms of scheduling television commercials. The company's sales people and retailers may give the test product special consideration—care that the product would not have in the national market. Competitors may sabotage the test, usually deliberately, but sometimes by accident. Take Procter & Gamble and its soft Duncan Hines brand cookies as an example.

Frito-Lay, with Grandma's Rich'n Chewy brand, was the first soft cookie to reach the test market stage, although P&G had been working on soft cookie technology for years. In the Frito-Lay test, Grandma's quickly captured one-fifth of the Kansas City, Missouri, cookie market, far exceeding Frito-Lay's most optimistic projections. What happened?

A Frito-Lay competitor (not, ironically, P&G), concerned about Grandma's threat, had sent a research company to Kansas City to learn why people buy the product. Because consumers had to eat the cookies to talk about them knowledgeably, the researchers bought cookies to consume during the interviews. In fact, they bought so many cookies they pushed Grandma's market share higher than it would have been without the research. By chance, P&G followed Frito-Lay into Kansas City with its Duncan Hines soft cookie brand, where it achieved the same highly skewed results for many of the same reasons. Indeed, the results were so positive, P&G decided to take the product national.

Unfortunately, by the time Procter & Gamble began the national rollout, the public had lost interest in packaged soft cookies (they were overprocessed and tasted it). P&G ultimately had to write off soft cookie factories in Tennessee and Florida.

The fourth major consideration is how to monitor the test results. One traditional measure is sales based on manufacturer shipments or warehouse withdrawals. However, inventory fluctuations can badly distort the sales pattern. Moreover, manufacturer shipments and warehouse withdrawals are not a sensitive measure of consumer response anyway. They reflect trade response to the manufacturer's marketing efforts, particularly promotion.

The company today may be able to obtain store audit data of actual sales figures, which are not sensitive to inventory fluctuations. Store audits also provide information about such variables as distribution, shelf facings, and in-store promotional activity, all factors that affect the product's sales.

Unfortunately, however, manufacturer shipments, warehouse withdrawals, and even consumer sales data tell us very little about the impact of the marketing program in building and maintaining a successful new product or service. What's required is intelligence based on a cross-section of prospective and actual buyers.

The company has to obtain measures such as consumer brand awareness, attitudes, trial, and repeat purchase from either consumer panels or surveys. Brand awareness and consumer attitudes also serve
as criteria for evaluating the marketing program and can help in interpreting sales data. As Aker and Day assert, "The most useful information obtained from consumers, however, is whether they bought the product at least once and whether they were satisfied with it and have either repurchased it or plan to." The Louis Rich division of Oscar Mayer (now owned by Philip Morris Companies, Inc.) used this information when it tested a line extension to its successful Breast of Turkey product, an item called Dark Roast of Turkey.

Because some people in the home office "were afraid we were going to shoot ourselves in the foot anyway," said Steve Lindbloom, a Louis Rich product manager, "we didn't want to go into Minneapolis or some other major market where we stood to lose some credibility with our trade if we bombed." Instead, the company tested Dark Roast of Turkey in Eau Claire, Wisconsin, and Midland-Odessa, Texas, both markets in which Breast of Turkey sold well. "We only had one media campaign, but we split cells to see what would happen if we didn't advertise at all in one of the test groups."

Virtually nothing happened in the cell without advertising. The other cell did generate a good trial rate, but repeat sales were unacceptable. "Initial trial was 18 percent," Lindbloom recalled, "but repeat was 4 percent—and we decided that we couldn't live with that." Louis Rich also discovered that the period between initial trial and the first repeat purchase was much longer than it had expected. "That gave us some projections of the size of the idea—and it was much smaller than our assumptions had led us to believe."

Earlier in his career, Kevin Clancy, one of this book's authors, worked on a "convertible cigarette," a seemingly ordinary filter cigarette with a breakthrough, magical property. The filter contained micro-encapsulated menthol that the smoker could activate by squeezing the filter. Early returns from test markets indicated the company was enjoying tremendous success; smokers appeared to love the idea because sales during the first eight weeks rose faster than the morning sun.

Unfortunately, sales fell almost as quickly. Once they tried the product, smokers came to hate it. Sometimes a light squeeze gave a blast of menthol that was like a fist in the throat; other times it took a karate chop to release any menthol taste at all. The product enjoyed virtually no repeat sales, a fact that a tracking study clearly revealed. Without such a tracking study, the manufacturer might have mistakenly believed that it had a clear winner and geared up accordingly.

Health and beauty aids manufacturers have always found it difficult to conduct traditional test market research, said Gail A. Lanzer, director of marketing research at Helene Curtis, Inc. Because drug and mass merchandise accounts represent half to two-thirds of the business in many health and beauty aids categories, the manufacturers need these retailers for their test market research. Yet many drug and mass merchandise accounts do not have scanners at the cash registers, so the manufacturer has to audit the sales. Many of these retailers are large chains, and their management's are often unwilling to permit a manufacturer to use a certain store in a certain market for an audit. Health and beauty aids marketers also compete in some categories where purchase cycles are very long. A long purchase cycle makes a test market difficult to conduct in some cases and difficult to read in others.

Test Marketing Trends

In recent years, General Foods and other packaged goods companies have been shifting much of their in-market testing activity to the newly available, small electronic markets. These are markets in which the retail stores have UPC scanner checkouts and individually targetable cable television so that the company can control what consumers see and record what products they buy.

Jane Brown, a consultant in new product development and test marketing to Kraft General Foods, noted that, on the surface, testing of both new and established products in 1990—the date of her research—looked somewhat different than it did in 1976. In 1976, General Foods conducted forty-two traditional in-market tests; forty-four tests in 1986; and forty in 1990. So the number of tests had not changed dramatically. In 1976, the traditional test market took one to three years; by 1990, the company's tests—both in-market and mini-market—took four months to two years.

In 1976 General Foods selected its test markets scrupulously to replicate the average environment for a category/brand, evaluated a fully developed business proposition, and used the results for brand go/no-go decisions, considering them fully projectable. In 1976, the company would consider a mini-market only as a precursor to a
full test market, since "we 'knew' it was not projectable," said Brown. She said that some of the environments in which General Foods tests today are smaller than those they considered "mini" in 1976. "The selection process is less rigid in terms of representativeness and more likely to be a function of availability and acceptability; i.e., category/brand development not less than 80% and enough of the characteristics to do the key analyses. We continue to evaluate a fully developed marketing plan and, despite issues of predictive accuracy, we have been making broad-scale decisions from these experiences."

Companies have not changed the type of data they collect and the key measures they use to evaluate it. What has changed since 1976 are the methods for capturing the data and disseminating messages. UPC scanners at the checkout, consumer data cards and individual hand-held scanners (to link individual consumers to specific product sales), and individually targetable cable television has improved test market execution and design capability. These have also led to increased experimentation with marketing variables for both new and established products.

The basics of an in-market test from a "need to know" basis have not changed. "We are still monitoring trial and repeat sales, reaction to consumer and trade promotions, looking for a fit against sourcing assumptions, market growth, etc.," said Brown. But the go/no-go decision has become far more difficult as companies learn more and more about market dynamics. To understand them adequately, the company must understand historical trends, market nuances, variable responsiveness, and more.

The Trouble with Traditional Test Markets

Test markets are fraught with problems. Often the company selects a test market because it's easy to manage, or because a retailer in the market will cooperate with the test, and not because the market represents the target the company in fact wants to reach.

Traditional test market research has four major defects:

1. **Traditional test markets are expensive.** They can cost as little as $1 million, but typically run more. "Add up the research, the media, the effort throughout the organization to control it, check the test market, and the costs are hard to swallow," said Gail Lanznar from Helene Curtis.

2. **Traditional test markets take too long.** Waiting a year, eighteen months, or two years for results is simply not competitive in an environment where the pace of change has picked up as much as it has in many product categories.

3. **Traditional test markets give away ideas.** Marketers routinely gripe, "They're not secure, and as a result we're giving our competitors free marketing intelligence."

4. **Competitors can sabotage the results.** Even modest efforts by competitors can spoil the company's ability to read the test market results. Competitors have sabotaged tests by having their sales people pull the new products off retail shelves, turn them sideways, or move them to other shelves where shoppers will not notice them. (Fortunately, what competitors do in a test market to affect a brand's performance is not what they're likely to do in the national market, which is too big and unmanageable.)

As a more extreme example—but by no means unique—take the experience we had with what we've called elsewhere "The Case of the Annihilated Brand."

This food product's advertising and promotion budget, projected nationally, was $70 million. The company assumed that the competition, faced with the prospect of battling a new brand, might increase its advertising and promotional spending by as much as 80 percent over the base period. Using this assumption, a model was employed that forecast the new brand would obtain a 3.6 percent market share by the end of Year One.

The company introduced the brand in test markets—where the brand failed miserably. By the end of Year One, the brand's market share was 0.7 percent—a clear disaster. Was the research and the modeling flawed? Didn't the company support the brand adequately with advertising and promotion?

An autopsy revealed that the company's advertising and promotional spending were in line with what it had said it would spend. The competition's spending, however, had not increased 80 percent in the test markets. It had increased 630 percent. Projected nationally, this amounted to $1 billion. In effect, the competitive program "gave away" the existing product to regular buyers at a
heavily discounted price. Consumers wisely responded by stocking up, which took them out of the market for months. The move overwhelmed and annihilated the new brand.

Of course the competitor who increased its spending 630 percent could not have sustained the expense for very long and not on a national scale. At that rate, the competition was losing money—certainly in the test markets. But this was not a national introduction, and the management obviously felt it was worthwhile to lose money in the test markets to prevent the new brand from establishing itself. As Frank Brod of Coca-Cola Foods said in Chapter 1, “Minute Maid slaughtered Citrus Hill on the beaches. They kept the product off the shelves by loading the pipeline with Minute Maid.”

5. Traditional test markets usually fail to tell marketers what they need to know to achieve success. While a product failing in a test market isn't as painful as failing in the national rollout, it's often difficult to determine why. Was the problem with the way the company executed the idea or was the idea simply too small? Was the problem with the marketing program or with the competitive response? Typically, marketing managers test too few plans before introducing a new (or re-positioned) product into the real world. Most do not even realize how many plans they could test. Consider “The Case of the Escalating Sevens.”

Assume we are managing the introduction of a new soft drink, Zippy Cola, a product formulated with high levels of caffeine, sugar, and carbonation. Assume further that we have to make a dozen key marketing decisions and that we have only seven choices for each—seven target groups, seven positionings, seven advertising executions, seven price levels, and so on. The first decision we have to make is the market target. Here are our seven choices:

1. Women 18-to-49-years-old, the all-time most popular target for soft drinks.
2. Health freaks, the granola and oat-bran crowd.
3. YUMMIES, a hot new target. Young, Urban, Moneyed, and Single. These are the yuppies with high school teachers excluded and $100-an-hour plumbers included.
4. Friends of Ponce de Leon, the people looking for the fountain of youth, virtually all of whom belong to a health club.
5. Kids 6-to-12-years-old.
6. Folks who need a sugar and caffeine fix regularly.
7. Aging athletes, the people who relive their days as a high school football player or tennis star. This group is heavily represented by middle-aged men.

If there are a dozen ingredients in the marketing mix and exactly seven options for each, how many different plans might a marketing manager develop?

Not twelve times seven, but seven to the twelfth power: thirteen billion, eight hundred forty-one million.

Given such a staggering number of possibilities, how do modern marketing managers handle the problem? Most companies do a little research—six to eight focus groups, some concept, product, and copy testing. And then management uses its experience and best judgment to pick the best options from the different possibilities and puts together the best plan it can. This is the plan it takes into the traditional test market. But what are the odds of picking the best plan out of fourteen billion when the company does so little research? Not even one in a billion.

Our example is based on the assumption of a dozen marketing mix ingredients and exactly seven options for each. In the real world there may be many more than this or fewer, but our point is the same. Managers seem to select the marketing plans they take into test markets impulsively, and as a result marketplace failure abounds. Worse, little information flows back to the manager on how the marketing program could be improved. If a company actually tested three or thirty or three hundred of the plans it might develop, managers would have a wonderful basis to discover the factors that led to success or failure. Yet, traditional test marketing is simply too complicated and too expensive to permit such an analysis.

Managers are not the only people unhappy with traditional test marketing. Retailers are not happy either. Ronald Frost of the Fleming Companies, the country’s largest food wholesaler, has been quoted as saying, “We are in effect paying the cost of market research. They are putting it out there and seeing what happens. We don’t feel that’s an appropriate approach to market research.”
As a way to deal with the situation, some wholesalers and retailers are introducing a "failure fee" in addition to the slotting allowances, the fees manufacturers pay to retailers to gain shelf space for new items. The manufacturers pay the failure fee if a new item does not sell as well as agreed upon. It is, in a sense, "a guarantee of sorts to reduce the retailer’s financial risk" while increasing the manufacturer’s.

A Brief for Simulated Test Marketing

What can marketers do to improve the likelihood of new product success in an age of promotion and unprecedented competitive response levels? How can companies anticipate defensive competitive response and develop and test offensive strategies designed to overwhelm those defenses? And how can companies do this without costly, time-consuming real-world test marketing?

A well-done simulated test market reduces risks that include not only lost marketing and sales dollars but also capital—the expense of installing production lines or building a new factory to manufacture the product. Why would a company spend $2 million or $3 million and wait a year and half to learn of a failure about which it can do little when it can spend $100,000 or $150,000 and take three to six months to learn how to fix any problems?

A simulated test market study increases efficiency. If a company has, say, three new-product development projects underway, and one seems to offer more volume and greater margins, sagacious management would promote that project rather than the others. The STM can indicate the project offering the greatest return. An STM can also optimize the company’s marketing efficiency in a new product it does go ahead with—to see the effect, say, of shifting a budgeted $1 million from television advertising to a coupon or vice versa. (We discuss optimization in detail in Chapter 11.)

A simulated test market study maintains security. As soon as a company puts a product into a real-world test, everyone who cares knows about it, starting with the competition’s sales people. Several years ago, Procter & Gamble began testing a ready-to-spread Duncan Hines frosting. General Mills saw the test and rushed its own Betty Crocker brand introduction, which now dominates the category.

A simulated test market can save the company time. The STM can give you results in three to six months where you may have to wait more than a year for the same results from an in-market test.

Time, as an element of competitive advantage, is only beginning to gain currency. But as a strategic weapon, time is the equivalent of money, productivity, quality, even innovation. George Stalk, Jr., a vice president of the Boston Consulting Group, wrote in Competing Against Time, "The ways leading companies manage time—in production, in new product development and introduction, in sales and distribution—represent the most powerful new sources of competitive advantage. While time is a basic performance variable, management seldom monitors its consumption explicitly—almost never with the same precision accorded sales and costs. Yet time is a more critical competitive yardstick than traditional financial measurements."

A company that builds its strategy on flexible manufacturing and rapid-response systems is a more powerful competitor than one with a traditional strategy based on low wages or manufacturing cost efficiencies. "These older, cost-based strategies require managers to do whatever is necessary to drive down costs," Stalk wrote, "move production to or source from a low-wage country; build new facilities or consolidate old plants to gain economies of scale." These all do reduce costs, but at the expense of responsiveness—and in the 1990s, customers will be more interested in response than in price.

Time-based marketing allows companies to serve key customer needs quickly, which in turn creates more value. In this era, the total time required to produce a product or service—not cost—defines a firm’s competitive advantage. "Early adopters report that actions modeled on just-in-time—simplified flows, waste reduction, reduced setup times and batch sizes—can also dramatically reduce time in product development, engineering, and customer service," says Dr. Joseph Blackburn, a professor at the business school of Vanderbilt University and a long-time friend and consultant to the authors. "Firms able to achieve faster response times have reported growth rates over three times the industry average and double the profitability. Thus the payoff is market dominance."

Today’s better simulated test markets capture every important component in the marketing mix and assess the effect of any plan
on product awareness, trial, repeat rates, market share, profitability, and more. These STMs test any plan the marketer wants to consider—even a competitor’s. The marketer simply enters the plan into the computer and the model forecasts what is likely to occur month by month in the real world.

Some simulated test marketing methodologies are even smart enough to help recommend a plan and we have never seen a plan a sophisticated STM recommends that does not beat the one submitted by the product manager. Sometimes the margin is modest; sometimes the difference is overwhelming.

Simulated test marketing is the single best validated tool in all of marketing research. For a new packaged goods, the better STMs can forecast what will happen in the real world, plus or minus 15 percent.

This is not to say, however, there are no cases when an STM result goes awry. Perhaps the biggest failures come about because the assumptions on which the model made its forecast were flawed. If a company estimates a distribution level of 90 percent and obtains only 60 percent, the volume forecast can be off substantially because, in some product categories, distribution corresponds almost a one-to-one with volume.

But not only may the assumptions be mistaken, the market’s dynamics may change between the STM and the actual test market. The company may have a new competitor, one it did not know existed when it began the simulated test market research.

Sometimes the company’s commitment changes between the STM and the real-world market test—and we believe a company should do a real-world test, but only after a STM. Or, to put it another way, in most simulated test markets, companies assume adequate marketing support, support that may disappear by the time the firm begins the product’s national introduction. For a simulated test market study it’s easy to say that you’re going to spend $24 million on advertising because no one has to write a check. It’s something else to spend real money.

Discrepancies also arise between the simulated test market performance and the actual test market because the real world is messier than a STM. But discrepancies also arise between the test market and the national performance. Companies routinely obtain test market distribution levels that are much higher than they ever see again because the sales force is excited by the new product and the sales people work harder than usual. This sensitivity to the product’s success brings results the company never repeats.

But suppose the real-world test results are significantly worse than the simulated test market research results. We ask the client, “What’s happening in the test market? What are the shelf facings . . . the distribution? What is the trade activity . . . consumer promotion . . . your share of voice? What is the competition doing?” With these new inputs, the STM can virtually always match what’s going on in the market.

At that point we can ask different questions: “Given what’s happening, is there anything we can do, anything we can learn from the simulation that would produce a better plan? Given that the competition has increased its promotional spending in our test markets by 630 percent, can we add markets until it becomes just too expensive for them to continue?”

Today the goal of simulated test marketing research is not to obtain a simple volume forecast. The objective is to provide diagnostic insights that will help improve the likelihood of product success. Telling marketers they will obtain a 5 percent share or a 10 percent share doesn’t satisfy them. They want insight from the study that will help them build plans with an even lower risk of failure.

A good STM will tell you not only how you’re doing but what to do differently. A sophisticated decision support system combines simulated test marketing with mathematical modeling of the marketing mix. Such a system goes beyond forecasting first-year volume potential to providing insights into improving the advertising, the concept, the product and packaging, and the marketing plan itself.

Marketers can ask a state-of-the-science STM model to evaluate every ingredient in the marketing plan in terms of effects on sales or profits or both. The model will run hundreds—in some cases thousands—of simulations to identify those factors that contribute most to marketing success.

But what is the source of simulated test market research? How did this methodology get its start and what is its status today?
A Short History of Simulated Test Marketing Research

Given the problems of cost, time, and security connected with traditional in-market tests and the development of the computer, which allowed people to manipulate vast amounts of data inexpensively, researchers began looking into simulated test marketing more than twenty years ago. Many of the marketing research ideas (such as research design and measurement issues) that were later incorporated into simulated test markets were initially developed in the mid- to late-1960s.

At the beginning, these efforts followed two separate paths that eventually converged. We call the first the "mathematical modeling" path, the second the "laboratory experiment" path. The first started with the idea that researchers could take historical data—advertising and promotional spending, distribution, market share, and much more—and build a series of equations that would forecast new product sales.

The second started with the idea that if the researchers could "simulate" experimentally in a laboratory setting the process by which consumers learn about and buy a new product, it would be possible to project the real-world results from such an experiment.

By the 1970s, simulated test marketing research models had become a major force in the marketplace. The 1980s saw a number of major refinements in the STM procedures themselves and saw literally thousands of applications. During this decade, the major STMs underwent quite a bit of evolution and convergence—that is, they
are far more similar today than they were a generation ago—and now may resemble their original specifications only slightly.

The Origins of Mathematically Modeling New Products: DEMON

In 1966 Dr. David Learner, who was research director of BBDO Advertising, and James DeVoe, an associate research director at BBDO, and two Carnegie-Mellon professors, Abraham Charnes and William Cooper, published a pioneering paper titled “DEMON: Decision Mapping via Optimal GO-NO Networks—A Model for Marketing New Products.” They described a very sophisticated mathematical model of the new product marketing process that BBDO was testing with several clients at the time.

For the first time the DEMON model examined very complex relationships including those between advertising spending and consumer awareness and between promotion and new product trial. With DEMON, a researcher could forecast brand awareness, trial, repeat purchase, and sales using techniques that were the state of the science in the new and emerging discipline of management science.

The problem with the DEMON model was that it could estimate none of the other necessary market parameters such as trial and repeat purchases prior to an actual test market. Typically in a DEMON consulting engagement, a marketer would introduce into markets a new product; collect data on awareness, trial, repeat, usage rate, and other factors at a three or more points in time; and then use this data to estimate the parameters of the model. DEMON would then tell the company how the product would do at the end of the year.

This meant a marketer could use a mathematical model to make national projections from the test market experience and could, theoretically, use DEMON to improve the marketing plan before going national—a revolutionary concept at the time.

After three years of experimentation, which was accompanied by growing excitement in the industry concerning this new technology, the model fell into disuse. It was simply too complicated and too sophisticated. There were too many input parameters, the computing technology at the time was slow, and management found it difficult to understand what DEMON could do for them.

DEMON’s future was not helped when David Learner, its developer and champion, left BBDO to become president of a high-technology firm in Pittsburgh.

Edward I. Brody, a senior vice president at BBDO New York, points out that advertising agencies such as Leo Burnett, DDB Needham, and NW Ayer, as well as BBDO, were also pioneering the development of new product forecasting models during the 1960s. Two young management scientists at BBDO, Drs. Larry Light and Lewis Pringle, picked up Learner’s work and took the best of DEMON to create a newer, simpler, stochastic model of the new product introduction process, a model called NEWS, an acronym for New Product Early Warning System. NEWS, however, like its predecessor was also limited by its inability to make valid predictions before an actual new product test market experience.

BBDO developed NEWS as an independent model designed to provide an accurate and easily understood input to a marketing manager’s new product decisions. Companies initially used NEWS only with early test market data (typically obtained from the first two or three months of an in-market test) to provide a forecast and certain diagnostic information for the remainder of the test market period. This became NEWS/Market, and the agency used it exclusively until years later when BBDO’s Pringle and Brody had accumulated sufficient validation to develop another version of the model called NEWS/Planner. The only difference between the two versions is that the input for NEWS/Planner came out of consumer research the company conducted before introducing the product into test market.

A Competing Vision: The Laboratory Test Market

Learner and his associates were barely aware at the time of what Yankelovich Skelly & White were doing to create the Laboratory Test Market that YSW introduced in 1968—an example of independent, almost simultaneous inventions. The Laboratory Test Market measured consumer trial in a laboratory environment by making prospects aware of the new product through advertising.

Yankelovich Skelly & White exposed a group of about 500 consumers to advertising for a new product and its competitors. They then brought these people into a store and gave them an opportuni-
ty to buy anything they wanted, but clearly the store's focal point was the new product. People who bought the product took it home and used it. Later (the time varying by product and purchase cycle), YSW researchers phoned the consumers who had bought the product, asked their reactions, and asked if they'd like to re-order. Their responses gave YSW the probability that a consumer would re-purchase the product if he or she found it in a store.

Based on the answers to these and other questions, Yankelovich Skelly & White had an estimate of trial (or, more technically, an estimate of awareness-to-trial conversion given distribution), an estimate of trial-to-repeat conversion, and the usage rate. If consumers were to purchase the new product just as frequently as other products in the category, YSW would assign a usage rate index of 1. On the other hand, if heavy users were going to buy this product or if consumers were going to buy it more frequently than other brands in the category, its usage index would be greater than 1—1.1, 1.2, 1.3, or more.

The Yankelovich Skelly & White researchers would then apply a factor they called "clout." YSW used this variable to account for the effects of advertising spending, sampling, couponing, and the manufacturer's overall power. A small, obscure, weak manufacturer had less clout than a large, well-known, powerful corporation. A YSW committee based its clout figure on historical experience, and these estimates—and they were simply estimates—ranged from 0 to 100.

To develop a final forecast, Yankelovich Skelly & White would take the consumer trial figure, multiply it by the client's estimated distribution at the end of the year, multiply that by the estimated clout, multiply that by repeat purchase, multiply that by the index of usage rate, and the final result was an estimate of what YSW called "on-going market share" in units.

As an example: say that 40 percent of the prospects bought the new product in the laboratory store. Yankelovich Skelly & White, recognizing that people's behavior in the laboratory was generally overstated, developed norms by which they reduced the 40 percent to a number the firm felt more closely reflected what people would do in the world. Assume that in this case the researchers multiplied the 40 percent by .75, so the trial figure becomes 30 percent.

The 30 percent was affected by product distribution, so assume the manufacturer estimated that it would have 67 percent distribution at the end of the year. Multiply the 30 percent by .67 to obtain 20 percent. YSW now multiplied the 20 percent by the clout factor. If we assume a clout of 0.5, the trial figure becomes 10 percent.

YSW further reduced the trial rate by repeat purchase, and let's assume that half the people who originally bought the product will buy it again.

Multiplying the 10 percent by 0.5 reduces the repeat figure to 5 percent. Finally, assume that the people who were buying this product were either light users—80 percent of average—or were going to be using this product to fill 80 percent of the normal requirements. In either event, YSW multiplied the 5 percent by 0.8, which yields a 4 percent year-end share of market.

The system, though primitive by today's standards, worked remarkably well. Over its first ten years, the Laboratory Test Market proved to be accurate plus or minus 10 to 15 percent about 90 percent of the time in forecasting a new product's first year sales.

The Pillsbury School of Simulated Test Marketing

The work that ultimately led to today's ESP and BASES simulated test marketing systems began at the Pillsbury Company in the early 1960s. Gerald J. Eskin, who became one of the founders of IRI but was then at Pillsbury, has said that his marketing research mentor, Dudley Ruch (in our view, one of the most brilliant, innovative, and supportive people in the marketing research profession), asked him to work on the following problem: Suppose the company has three to six months of data from a test market. Suppose the data included some diagnostic information on purchasing patterns—who's buying the new product and how many times they buy it. From this information, said Ruch, forecast the product's first year sales and the likelihood that the product will grow or decline thereafter.

Eskin says that he started with trial/repeat modeling. He went back to work that Joseph Woodlock had done at MRCA in the 1950s with Louis Foutt. "I took the Foutt Woodlock Model and expanded it," says Eskin. "I studied each repeat level separately. First repeat. Second repeat. Third repeat, etc. I also looked at repeat in a new way, one that took account of the amount of time that a person had to make a repeat purchase. These, we called 'true repeat curves.'"
These curves described both the probability that an individual would repurchase a product and the distribution of time until the next purchase. The Pillsbury researchers called the finished model "PanPro" for Panel Projection.

In one sense, as Eskin freely acknowledges, PanPro was a kind of sales accounting model. It described how people go through the process of first buying a new product, waiting a time, deciding to buy the product again, then perhaps waiting a time again, and if they decide to repurchase, making a third purchase, and so forth. PanPro was not a model that would forecast the outcome of a test market. It was not a pretest market simulator.

When Gerald Eskin finished the PanPro project at Pillsbury, John Malec visited Pillsbury and asked Eskin if he would be interested in developing a model that would work prior to the test market—a simulator. At the time Malec was working at the NPD Group, Port Washington, New York. Eskin reports that he did develop such a model, but not by building it from scratch. Rather, he collected concept test data and product test data for a wide range of products and tied these measures into the PanPro model, which produced the ESP (for Estimating Sales Potential) model. The NPD Group has been marketing the ESPN model since 1975. The company introduced versions of the model designed specifically for line extensions and for product re-stagings in 1980 and 1983, respectively, and the model is known now as Simulator ESP.

Eskin says that shortly after ESP was developed, Lynn Y. S. Lin, who was also working at Pillsbury, inspired by the same Dudley Ruch, developed an interest in simulated test marketing research. "Lynn was exposed to my early work at Pillsbury and had knowledge of the PanPro model and the NPD work for that matter," Eskin reports. After leaving Pillsbury to go to Burke Marketing Research, in Covington, Kentucky, Lynn developed a simulator model along lines similar to the ESP model. This became the BASES model, which Burke introduced in 1978, a model that went on to become the most widely used laboratory simulation model in the world.

It's interesting to note that BASES became successful despite its modeling simplicity and attitudinal rather than behavioral foundations. In the Yankelovich Laboratory Test Market, as we've said, consumers were exposed to advertising for the new and for competitive products in a laboratory environment; they were brought into a store where they used their own money to buy the product; and after they tried it at home, they were asked if they wanted to re-order it. The BASES approach simply called for presenting the new product in concept form to consumers and asking them how likely they would be to buy. Their answers were on a five-point purchase probability scale—from "definitely would buy" to "definitely would not buy." BASES then multiplied each point on the scale by a conversion weight (for example, the "definitely would buy" scores might have been hit by a weight score of 90 percent and the "probably would buy" by, say, 75 percent.

Following a period of in-home use, BASES gave consumers the same rating scale and asked them how likely they would be to buy the product again. BASES then applied a second set of weights. BASES then multiplied a string of figures to estimate the volume sales at the end of the first year following launch: the client-provided year-end awareness figure, times the projected trial number, times distribution, times Nielsen ACV distribution estimates, times the forecast repeat purchase figure, times the purchase rate. For more than fifteen years this simple, but very successful, model has demonstrated a satisfactory level of validity and has gone on to become the market leader in terms of number of new products tested and sold.

The Academically Grounded STM Systems

Management of Decision Systems, Inc. (MDS) introduced the ASSESSOR model in 1973 after initial development work by Professors Alvin J. Silk and Glen L. Urban of the Sloan School at MIT. Unlike all previous laboratory methods, ASSESSOR received an academic imprimatur when two seminal papers based on the model were published in the Journal of Marketing Research and presented at numerous conferences.

The ASSESSOR model conceptualized awareness and based the awareness figure on estimates developed between the client and the client's ad agency. The ASSESSOR researchers measured trial in the laboratory setting, much like Yankelovich, but instead of measuring trial by buying behavior—the Yankelovich approach—ASSESSOR researchers gave prospects vouchers they could use to
buy any product they wanted in a model store; the coupon-induced trial behavior coupled with a sophisticated constant-sum-based preference model were the ASSESSOR analogs to Yankelovich laboratory buying behavior and the BASES five-point purchase probability scale.

Like its predecessor, the Laboratory Test Market, the ASSESSOR model called for contacting “buyers” of the test product a few weeks after the purchase to gather information about attitudes toward the new product and re-purchase intentions and plans. The ASSESSOR model used a constant sum model to estimate what proportion of on-going sales a brand could obtain. They told prospects, “If you could divide eleven chips among the leading brands in the category, how many chips would you give to the different brands?” In other words, they employed two different approaches for estimating trial and repeat, one was behavioral, the other was attitudinal.

As a consequence, ASSESSOR had an awareness-to-trial conversion estimate, just as Yankelovich had its clout-to-trial conversion estimate. ASSESSOR multiplied that figure by on-going distribution, multiplied that by an estimate of repeat purchase and by a usage rate, and that gave them an on-going share estimate.

Any comparison of the leading laboratory STMs circa the late 1970s cannot help but give high marks to the Yankelovich Laboratory Test Market, in contrast to most other models because (a) it was based on actual spending behavior as opposed to coupon redemption methodologies or attitudinal responses (or both), and (b) Yankelovich Skelly & White consultants had more years of experience in estimating clout than ASSESSOR or BASES researchers had in working with the awareness numbers provided by clients and their ad agencies.

In fact, clients and their ad agencies almost always inflated the awareness numbers by a considerable margin. Given that a linear relationship exists in all these models between where they start and where they end, to the extent that someone exaggerates an awareness estimate, every other figure will also be exaggerated—most significantly the year-end share estimate. Marketers praise the modeling intricacies of ASSESSOR, however, and the simplicity and relatively low cost of a BASES study, thus explaining in part the strong success of these methods.

In 1979, Malec and Eskin started Information Resources Inc. (IRI), which acquired MDS in the mid-1980s. “As you might expect,” says Eskin, “we couldn’t resist the temptation to try to improve on our earlier work. But as you might also expect, the ASSESSOR model also has some of the flavor of those earlier PanPro-based models.” Assessor’s preference structure models were merged with IRI’s FASTRAC in-store scanning system database; and the model became known as ASSESSOR-FIT in 1985.

The M/A/R/C Group of Irving, Texas, bought the rights to the ASSESSOR model in July 1989. Before that M/A/R/C had ENTRO, its own simulated test marketing system, with which it had been successful. M/A/R/C established a new subsidiary company, MACRO Strategies Inc., to provide marketing consulting and modeling services, including simulated test market research studies. MACRO Strategies merged the ASSESSOR and ENTRO models, which it currently offers as MACRO ASSESSOR. We, however, will call the model simply ASSESSOR.

The Development of DESIGNOR

In the early 1970s, Jacques Blanchard was sent to MIT’s Sloan School of Management by a major European management consulting firm. The firm wanted him to develop a market modeling business in Europe, and he worked as an MBA student of Glen Urban’s. Blanchard worked with Urban on the European implementation of SPRINTER (an early model similar to DEMON in design that made predictions on the basis of early in-market tests) and on PERCEPTOR, which was introduced in Europe in 1973.

PERCEPTOR is a tool to understand consumer preferences for different products as a function of their competitive positioning and buyer perceptions. It relates preferences to buying behavior. Blanchard says in fact that some of the findings of ASSESSOR came from earlier work done on PERCEPTOR. “There already was a trial and repeat mechanism in PERCEPTOR which was a new concept for ASSESSOR. We already had a lot of experience in the evaluation of concepts and how people react to perception before launch, and we also had a lot of work using market response modeling on in-market data for new products, both in the US and in Europe.”
The ASSESSOR was developed, and Blanchard and his associates tested ASSESSOR's micro-structure in Europe before it was released there in 1975 (and in Japan shortly afterward). At the time they were using the PERCEPTOR technology to evaluate concept-use placement tests and make projections. Blanchard did some micro-structure validation of the trial and repeat model in ASSESSOR that they had used in developing PERCEPTOR. In 1979, his firm, Novaction, headquartered in Paris, introduced the PERCEPTOR/Concept test system based on ASSESSOR, and in 1983 introduced several new modules covering line extensions, relaunches, and price elasticity.

Novaction introduced DESIGNOR in 1986, in which the emphasis is heavily on diagnosis and optimization, rather than simply volume forecasting. DESIGNOR was a combination of PERCEPTOR, ASSESSOR, and several new modeling tools, including price optimization and an awareness forecasting subsystem, designed to make forecasts in a unique way not available through MDS in the United States.

Today's DESIGNOR forecasts trial using a combination of the voucher methodology and the sophisticated preference model that made ASSESSOR famous, integrated with repeat purchase, usage rate, and awareness forecasting approaches that are entirely its own. These tools coupled with a third forecasting methodology, an analog model, which forecasts new product performance based on the IDQV (impact, differentiation, quality, and value) of the new product compared to the IDQV of other products with similar market structures whose sales performance is known, provide DESIGNOR with a powerful arsenal of modeling tools and make it a formidable competitor in the STM industry.

At the Interstices of Mathematical Modeling and LTM: The Origin of LITMUS

In 1977 Florence Skelly, one of the Yankelovich Skelly & White principals and a pioneer in the field of marketing research, met Kevin Clancy, who she knew was interested in mathematical models. She also knew that he had worked on the NEWS model at BBDO Advertising. Concerned that ASSESSOR was having a negative impact on Laboratory Test Market sales, Skelly asked Clancy if he could combine the Laboratory Test Market method and databases with a mathematical model of the new product process to improve the Laboratory Test Market system's capabilities. Clancy thought he could, and he and Professor Joseph Blackburn spent the next year working on a model they eventually called LITMUS. Robert Shulman, a vice president and sales manager for LTM at the time, became their in-house consultant and most enthusiastic supporter, constantly pressing them to improve the capabilities of the evolving model.

To test the LITMUS model, Skelly and Robert Goldberg, a Yankelovich Skelly & White new products guru who had been working on the Laboratory Test Market for more than ten years, gave Clancy and Blackburn twenty marketing plans for which YSW already knew the real-world results. They asked Clancy and Blackburn to run a forecast through the LITMUS model that they could compare to the actual national introduction.

At the meeting to discuss the results, each group had twenty envelopes. Clancy's envelopes contained the forecast of what the product would do based on the marketing plan input, laboratory response, and LITMUS's calculations; the Laboratory Test Market veterans' envelopes held the products' actual results. The veterans expected major differences between the forecasts and the actual results because they found it difficult to believe that a mathematical model could equal the intelligence and expertise of a group that had been making new product forecasts for years.

In fact, the LITMUS model's forecasts were virtually identical to the actual results in seventeen of the twenty cases. The results were so close that Clancy and Blackburn decided to write up the results for publication and Yankelovich began to market the program in 1981.

One of the three cases in which LITMUS produced a result that was quite different the real world experience was for a new peanut chip (a peanut equivalent of the chocolate chip, an essential ingredient in chocolate chip cookies). LITMUS factored its trial estimates by expected Nielsen All Commodity Volume (ACV) estimates and produced a sales forecast which was just about half what the manufacturer actually achieved in the real world.

In analyzing the differences between the model and the actual
experience, it was found that the model didn’t take into account the fact that many housewives, pushing their carts down supermarket aisles, would reach out to grab a package of chocolate chips to quiet a restless child. They would see the new peanut chip product beside chocolate chips on the shelf and take it for the novelty or—in some cases—by mistake. As a result, the usual effect of product distribution was considerably different (and greater) for this product than for most others. Real-world sales were terrific.

LITMUS had a number of properties that were unique at the time. Its submodel for forecasting awareness, as an illustration, contained thirteen different determinants of what awareness should be for a new product—including advertising impact and gross rating points by time period, media impact, and forgetting coefficients. This differentiated LITMUS from the original LTM, which didn’t have a formal awareness function at all, and from the ASSESSOR model and other simulated test marketing models introduced later, which called for advertising agencies and clients to estimate awareness.

In addition to the awareness submodel, the evolving LITMUS enjoyed some other technological breakthroughs, among them that the model did not assume (as all models did at the time) a linear relationship between distribution and sales—an innovation inspired by the peanut chip case. Research had discovered that the more involved consumers are in the product category, the more likely they are to shop in several stores to find the product. One can imagine a situation where consumers are so involved that they will go from store to store to find a new entry until they do find it. Hence, in those categories, for those consumers, a 10 percent distribution level might act in the same way that an 80 or 90 percent distribution level might perform for a very low involvement product. LITMUS took this into account by estimating each consumer’s involvement level and correcting distribution by that knowledge to estimate the real distribution effect.

Among the many differences between LITMUS and other models at the time, we should mention a third. For the first time in an STM model, LITMUS took into account the product’s purchase cycle, differentiating between a product that a consumer might buy once or twice a year and one a consumer would buy weekly. Because LITMUS was able to provide weekly or monthly estimated sales and accumulate them over time, the model needed to know the amount of advertising and level of distribution in every month. Other models, ignoring these issues altogether, could not differentiate between time periods when a corporation might have very high advertising spending levels and no distribution or periods when the firm had high distribution and no advertising.

The methods and models discussed here are, of course, not the only simulated test marketing approaches available today. The published academic literature discusses two other simulated test market research programs: Elrick & Lavidge’s Comp and BBDO Worldwide’s NEWS/Planner. In addition to the STM that the academic literature has examined, a variety of other simulated test market research services are available commercially: Leo Burnett Model by Leo Burnett (Chicago); and Adopter by Data Development Corp. (New York), Tele-Research’s Micromarket, ASI Market Research’s Purchase Action, Market Simulation Inc., and Robinson Associates’ Speedmark. The literature does not explain them in detail, however, and it is therefore impossible to know from the outside how these models work.

What has been observed over the passage of time is both the evolution and the convergence of all the different systems available. The systems have both been improved, in terms of their capabilities, and they have converged in the sense that they have become more similar to one another.

As an illustration, when Yankelovich Skelly & White launched LITMUS it was the first simulated test market methodology to model awareness. Although the LITMUS awareness subfunction is arguably the most sophisticated in the industry even today (and the only one that’s been examined by academics), other simulated test marketing services are forecasting awareness as well. ASSESSOR, BASES, and DESIGNOR all have their own awareness forecasting approaches, which appear to work quite well.

Another area in which LITMUS played a revolutionary role was in optimization technology, not only the formal optimization routines, which we describe later in this book, but in terms of the general use of this new and emerging technology to evaluate many different marketing plans prior to an actual test market introduction.
to discover which plan, or combination of plans, appears to work best. Today all of the services appear to be taking a similar approach to this problem such that there is now a blurring of the distinctions between the commercially available technologies.

Who Uses Simulated Test Market Research?

Given this history of the development of STM models, how are corporations actually using them? The Advertising Research Foundation has conducted two surveys to gather industry-wide data on simulated test market research. A 1988 study established the STM market size and trends, how and when manufacturers use STM, how valid they perceive the results, and the benefits and limitations. The 1990 study’s objectives were to gather information and set directions.

The 1988 study found that among the six research firms reporting (which represented 80 percent or more of all simulated test market research in America), there were an average of 680 tests in 1986/87, the latest figures available. The average test cost $45,000, and the total market for simulated test market research was approximately $30,500,000.

The survey found that the majority (64 percent) of simulated test market research in 1987 was for new products; line extensions accounted for a third of the tests, and the remaining 3 percent was for established brands. Further, the STMs were being conducted almost exclusively for packaged goods manufacturers. More than half (56 percent) were for food products, with health and beauty aids (21 percent) and household products (16 percent) making up most of the remainder. The “all other” category includes not only tests for such things as financial services and consumer durables, but includes many tests for packaged goods, such as pet foods and beverages.

In surveying marketers, the ARF found that a few large manufacturers seemed to account for a disproportional high share of STM activity. “The most active 20 percent of the 42 responses we received, accounted for about two-thirds of the STM’s conducted by this sample,” said Allan L. Baldinger, senior vice president and director of marketing research at the ARF. “In fact, I believe that the slowdown in overall activity since 1985 is largely due to a few manufacturers who are cutting back from their peak levels, whereas most smaller advertisers are still on an upward growth curve.”

What Companies Do with STM Results

The 1988 survey also looked at what manufacturers do with the simulated test market research. A company can do five major things, said Baldinger: “You can go national, if your proposition looks good enough, or if the fear of preemption is great. You can go to test market, and, given consistently positive results, then go national. You can go to test market, get poor results in test market and then discontinue the product. You can use the STM for product optimization, by identifying some fixable weaknesses in your concept or product and recycle the concept for further testing. Or, finally, you can use the STM to reduce your test marketing risk, by discontinuing poorer performing ideas.”

Exhibit 3–1 shows the results of almost 600 tests, in total and comparing companies that run many tests versus companies that use relatively fewer. Baldinger pointed out that the seven companies that use STM’s heavily—spending several hundred thousand dollars a year on such tests—“were much more likely to be using

Exhibit 3–1

<table>
<thead>
<tr>
<th>Results</th>
<th>Percent of All Tests</th>
<th>Percent Among Heavy Users</th>
<th>Light Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1: Go National</td>
<td>14</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>#2: Test Market, Then Go National</td>
<td>12</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>#3: Test Market, Then Discontinue</td>
<td>14</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>#4: Recycle</td>
<td>16</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>#5: Discontinue Project</td>
<td>45</td>
<td>54</td>
<td>25</td>
</tr>
<tr>
<td>(Base: Total Completed 1985-1987 Tests)</td>
<td>(589)</td>
<td>(381)</td>
<td>(208)</td>
</tr>
<tr>
<td>(Base: Companies)</td>
<td>(7)</td>
<td>(3)</td>
<td></td>
</tr>
</tbody>
</table>

Source: Advertising Research Foundation, “Trends and Issues in STM’s: Results of an ARF Pilot Project.”
STM's to screen out poorer performing ideas than were the selective users. This seems to suggest that the heavier users are using STM's to broaden their scope of new product investigation, to widen the mouth of the new product funnel.”

Results of Simulated Test Market Research

The survey also found that the lighter, more selective users, were much more likely to spend more money per test than the heavier users. Also, almost two-thirds of the tests conducted by the heavy users were based on concepts only, said Baldinger, while a similarly high proportion of lighter users' projects involved a product use phase.

Finally, this survey asked the manufacturers, based on the tests they conducted between 1985 and 1987, how many the market confirmed, how many generated higher results in the market than the research predicted, and how many generated lower results. “We gave them no definition of how close the final volume estimate had to be to the original estimate to be validated. We asked only for their perceptions of validity.”

According to the manufacturers, slightly more than half (52 percent) of the in-market results confirmed the simulated test market research; a few (8 percent) of the in-market results were higher than the STM predicted; and the rest (41 percent) were lower. More than half (55 percent) of these 42 manufacturers said the single most important benefit of STM research was risk reduction; a third said the most important benefit was the diagnostics/what if capability. Other benefits included confidentiality (mentioned by 14 percent), cost effective (10 percent), product optimization (7 percent), and speed (7 percent).

This 1988 questionnaire asked simulated test marketing's single most important limitation, and more (40 percent) manufacturers named validity, with a typical complaint that STM's don't “reflect highly complex situations accurately.” Although this may be true, the fact remains that simulated test marketing research is the single most validated tool in marketing research. Other limitations included analytical/diagnostic issues (mentioned by 31 percent of the manufacturers), methodological limitations, particularly when they involve estimating long-term repeat rates and products with long purchase cycles (mentioned by 24 percent); category appropriateness/experience (17 percent); while the rest (12 percent) were generally negative.

The study found that companies frequently use natural sell-in tests for new products, and seven in ten manufacturers use them to test line extensions. “This could easily be due to the fact that trade acceptance is a critical component of line extension tests,” said Baldinger, “and hence the need for a real-world reading, including the trade as a test vehicle.”

The study found that manufacturers seem to be using simulated test market models for new products (94 percent of the respondents cited this reason) and line extensions (78 percent). These respondents are not using STMs extensively for other purposes; for example, few said they use STMs to test advertising spending/weight or copy strategy (13 percent), consumer promotion (12 percent), trade promotion (11 percent), or other marketing issues (16 percent).

Although we have briefly described the differences in simulated test marketing models in this chapter, how do the major simulated test marketing systems compare in their methodologies and outputs? The next chapter covers them in depth.
How the Major Simulated Test Marketing Systems Compare*

All the major STM research models—ASSESSOR, BASES, DESIGNOR, LITMUS, and Simulator ESP—are similar enough that they exhibit some common strengths and weaknesses. One of the things that they do best is to quantify topics that were mostly qualitative before the models were invented. For example, concept tests and product tests now have more meaning than they had before the models existed. A product marketer now has more to work with than, say, a 3.2 score on a “probability of buying” scale. Today, through modeling, the computer can combine knowledge of the relationship between consumer self-reports of likely behavior with actual behavior and other data (such as retail distribution and marketing plans) to develop sales estimates with reasonable levels of predictive validity.

The models allow marketers to take more factors into consideration when making forecasts. The earliest models focused primarily on the product itself. Over time, researchers paid more and more attention to such marketing variables as distribution, advertising, trade and consumer promotion, and as a result models today include most marketing factors.

Basically, however, all simulated test marketing models have consumer inputs and marketing assumptions that the research company feeds into the “model.” The model uses sophisticated mathematics (we’ll see some in the next chapter) with judgment and logic checks to produce the output.

This chapter explores some of the features of simulated test market research methodology and discusses the marketing research procedures different companies use to collect the data that serves as STM input.

We focus on the five simulated test marketing research programs that are the most popular in the commercial marketplace—ASSESSOR, BASES, DESIGNOR, LITMUS, and Simulator ESP. We present some of the major features of these simulated test market research methodologies, emphasizing their similarities and differences. This discussion focuses on six major elements:

1. What information the research requires from the new product marketer.
2. How the research company designs the “shopping” environment.
3. How the company selects respondents.
4. How the company collects data.
5. How consumer reaction to the new brand is measured.
6. What diagnostic information and sales forecasts are provided.

We compare each simulated test market research methodology’s approach, and we focus on the specific details associated with each.

The ASSESSOR Model. According to information provided by the company, ASSESSOR has been applied to over 1,300 new product introductions. While the vast majority of ASSESSOR applications have been in consumer packaged-goods categories, the model has also been applied to consumer durables, restaurant and fast food services, and financial services.

ASSESSOR currently has five components or phases, designed to guide management at various product development stages: Phase I, Concept Evaluation and Screening; Phase II, Product Positioning (with volumetrics); Phase III, Positioning and Product Evaluation (respondent callbacks and volumetrics); Phase IV, Total Proposition Evaluation (with laboratory store); and Phase V, In-Market Evaluation (post launch or multiple callbacks, or both). Our analysis will focus on Phase IV.

The volumetric projections for the various phases are derived and converged from three independent models in a “check-and-balance” forecasting procedure. Discrepancies in the forecasts from any one module force analytical re-examination of the parameters used for all. ASSESSOR uses a behavioral decision (choice) model; a modified, attitudinal preference model; and a period-by-period (Markov) depth-of-repeat model.

Documentation for many different aspects of ASSESSOR’s original specification and measurement tools can be found in Silk and Urban (1978), Urban and Hauser (1980, pp. 393–417), Urban and Katz (1983), and Urban et al. (1983). Also, Shocker and Hall (1986) include a discussion of how the former ASSESSOR-FT model worked. Of all of the STMs available commercially, the initial details of ASSESSOR’s approach to simulated test marketing are documented most thoroughly in the published literature. This chapter, however, is based on the literature and on information supplied by David Lipson, Director of MACRO Strategies, Las Colinas, Texas.

The BASES Model. The BASES approach to STM modeling is one of several separate procedures used at different stages of the new product planning process. Like ASSESSOR, BASES has separate models for everything from concept evaluation to in-market testing. The most widely used model BASES III and it combines a concept test with an in-home use test with a proprietary, unpublished mathematical model for diagnosing and forecasting new product trial and repeat purchasing. We will refer to it simply as BASES. The research tells clients how well a product fulfilled triers’ expectations and the likelihood that triers will remain loyal customers. Since 1989, the BASES Group, a division of BBI Marketing Services Inc., Covington, Kentucky, has marketed the BASES group of procedures.

Although market share figures are not available, BASES appears to be the clear market leader in STM procedures. Company brochures indicate that the firm has conducted over 4,000 BASES studies in the United States, with additional studies overseas. Of these BASES studies, approximately 50 percent have applied to food products, about 30 percent have applied to health and beauty
aids and household products, and about 20 percent to other types of consumer products.

Unfortunately, the BASES procedures have not been well documented in the published academic literature; although an article by Lin et al. (1980) provides some sketchy information on how BASES data are collected. But precise information on measurement techniques and model specification used by BASES has not been made available. Review articles on a variety of STMs by Robinson (1981) and Shocker and Hall (1986) provide additional information on BASES and describe some of its features. Nonetheless, BASES remains more of a "black box" approach to simulated test marketing than any of the four other STMs we review here. Our information about BASES comes from company literature and discussions with our clients.

**The DESIGNOR Model.** Since its introduction in 1972, DESIGNOR has been applied to over 200 product and service categories in forty countries. According to company literature, more than 2,000 studies have been conducted in hundreds of product categories including fast-moving consumer goods, consumer durables, services, and pharmaceutical products.

DESIGNOR was developed from the ASSESSOR model, which, at the time, was a model to evaluate market share for a new product entering a well-defined category and which needed finished marketing mix materials (advertising, package design, promotion responses, and the like) in order to function. DESIGNOR evaluates volume sales for a new product and works equally well on a product that does not belong to a well-defined category. Marketing mix materials used in the test do not need to be in final form. DESIGNOR will estimate the volume potential of a new brand, a line extension, or a re-launch. The firm offers four main simulation services that address the major steps in the marketing process. DETECTOR is an entry strategy and brand mix evaluation tool to help assist decision makers with share/volume forecasts; PERCEPTOR obtains data on critical success factors in a market; and BRAND HEALTH CHECK optimizes the use of allocated marketing resources for both new and established brands.

This chapter is based on literature supplied by Novaction’s General Manager in the United States, Tony Hufflett, who is headquartered in South Natick, Massachusetts.

**The LITMUS Model.** The LITMUS procedure has evolved greatly since Yankelovich Skelly & White introduced the Laboratory Test Market, which contained some of the original ideas, in 1968. Since the first LITMUS introduction in 1981, procedures for simulated test market modeling have evolved substantially. Yankelovich Partners Inc., Westport, Connecticut, currently markets the LITMUS model. According to information provided by the firm, the LITMUS model and/or its predecessor LTM have been applied in approximately 3,000 studies of new product introductions. Approximately half of these studies have been conducted in food product categories, with the remainder applied to health and beauty aids, household products, consumer durables, consumer soft goods, consumer services, and financial services.

The documentation for the LITMUS model has been provided by Blackburn and Clancy (1980, 1982, 1983), and Blackburn, Carter, and Clancy (1984), and Blackburn, Clancy, and Wilson (1989, 1991). Additional information on LITMUS is included in Shocker and Hall’s review article (1986). Interested readers may consult articles discussing aspects of the LTM model (Robinson 1981; Yankelovich Skelly and White, Inc. 1981) and the NEWS model (Pringle, Wilson, and Brody 1982) to obtain additional perspectives on how LITMUS works. An on-going research and development program has been conducted on the LITMUS model since its introduction, and this book contains some of the unpublished research.

**The Simulator ESP Model.** The STM procedure currently known as Simulator ESP (for Estimating Sales Potential) has been marketed by the NPD Group, of Port Washington, New York, since 1975. The company introduced versions of the model designed specifically for line extensions and for product re-stagings in 1980 and 1983, respectively.

Documentation for the ESP approach to simulated test marketing can be found in Eskin’s original work on PanPro (Eskin 1973) and Eskin and Malec’s extension of PanPro to pre-test marketing situations (Eskin and Malec 1976). Narasimhan and Sen (1983) published a critique of the PanPro model (as well as of eight other test-market models); it provides a good analysis of some of the PanPro model’s features. Interested readers should also see an article
by Kalwani and Silk (1980), which presents some of the hypotheses on the structure of repeat buying used by Eskin (1973) as well as some of interesting aspects of the PanPro/ESP model parameter estimation process.

While the early work on PanPro and ESP appeared in the published literature, little on the ESP approach to simulated test marketing has been published since 1976. According to information obtained from the NPD Group in Port Washington, New York, well over 2,000 studies have been conducted using ESP and Simulator ESP. The ESP approach has been applied only to consumer packaged goods, with 65 percent of the applications conducted in food product categories.

Because the company has unique access to NPD/Nielsen Purchase Panel data bases, which it uses to estimate purchase cycles and buying rates for new products entering established categories (it recommends sales waves for innovative products), it argues that its estimates are more accurate than survey data. Further, the firm uses purchase panel data to update ESP model parameters to incorporate recent category sales trends; this keeps the model current and minimizes "model" bias. It also uses purchase panel data for on-going model development such as Line Extension ESP, Restage ESP, and Subline ESP.

**Typical Design of STM Research Methodologies**

It is important to understand the general approach that all simulated test market research procedures take. In many ways, most STMs are similar in their inputs, the approach they use for data collection, and their outputs. Reviewing the similarities between simulated test market research procedures will help clarify each procedure's specific details that we discuss later. Note that here we refer to the marketer of the new brand or service as the client and the marketing research organization that owns the simulated test market procedure and coordinates the STM methodology as the STM consulting firm.

Before the STM consulting firm can begin to collect data, the client must provide a stock of the test product. Preferably, the client has packaged the test product with the finished logos and graphics that it will use when it actually introduces the product into the market.

The client also must provide test product advertising, preferably in finished form. If necessary, however, the test product advertising can be in rough or even concept-board form. This is certainly the case for BASES, which is based on exposure to a concept, not to advertising. In addition, for ASSESSOR, DESIGNOR, LITMUS, and Simulator ESP, the client must provide advertising for the top competitors in the product category in which the new brand will compete.

If a client supplies finished advertising for competitors and rough advertising or concept boards for the test product, however, the difference may alert respondents to the fact that the STM consulting firm is interested primarily in reactions to the test product, thereby skewing the test results. For this reason, the simulated test market research should use finished (or rough) advertising for all brands so that respondents cannot easily guess which is the test product.

In addition to the test product, test product advertising, and competitive product advertising, the client must supply key information from the test product's marketing plan. A simulated test market study typically requires the following information for the test and competitive brands:

1. The market's size, stated in terms of the number of target consumers, households, dollars, or all three.
2. The advertising budget in dollars and gross rating points allocated by time period, e.g., monthly and media type.
3. Results of the test product advertising's copy tests. Generally, a client can use any one of the many copy testing techniques, as long as the STM consulting firm can compare the results to a norm or data base of advertising effectiveness relative to other brands in the same product category.
4. The types of promotional techniques planned and the budget for each.
5. The test product's price and the prices of the major competitive brands in the product category.
6. The expected distribution build over time, measured as Nielsen All-Commodity Volume (ACV).
7. The expected margin contribution as well as the total costs budgeted for marketing the new brand.
Design of the "Shopping" Environment

The laboratory environment that simulated test market research consulting firms use to collect data varies somewhat, depending upon the specific methodology. There seem to be three major types of "shopping" situations that consulting firms use to simulate a real test market.

1. People are exposed to advertising in a laboratory environment, either in a rented space in a shopping mall or in a traveling laboratory van, and then are given the opportunity to "buy" the test product and competitive product—and in some cases products in other categories as well—from a mock supermarket/drug store shelf, which is set up in the laboratory environment. One could think of this type of shopping environment as a mini 7-11.

2. Similar to #1 insofar as it involves a rented space in a mall or in a traveling van. However, there is no shopping experience per se. Researchers show people a concept board and ask how interested they are in buying the test product, typically on a five-point rating scale. The researchers then give consumers the new product to take home, so there's no buying involved at all.

3. Consumers are exposed to the new product in a facility adjoining a supermarket. They then go into the supermarket and actually buy the test product (or not).

Given the wide variety of types of shopping environments that STM consulting firms use, it seems there is a wide range of opinions on the "best" way to simulate the shopping experience that might occur in an actual market.

Respondent Selection. The STM consulting firms recruit participants at shopping malls, grocery stores, or by telephone. They sometimes screen potential respondents for product usage (or potential usage). In all cases, they collect basic demographic characteristics and use this data in defining the sample. To control for security and bias, they do not permit consumers who are employed by competitive organizations (for example, ad agency employees) to participate in an STM study. Average sample size ranges from a low of 250 to 300 respondents to a high of 1,500 to 2,000.

Data Collection Procedures. Major STM methodologies employ a wide range of data collection procedures, depending primarily upon the type of shopping environment the STM consulting firm uses. The typical STM data collection methodology is the following:

1. The consulting firm conducts an initial interview, using either a self-administered questionnaire or a personal interview, to measure brand awareness, brand attitudes and preferences, and product usage for the product category’s major brands.

2. The consulting firm then exposes respondents to advertising for the test product and some of the leading competitors. Typically, the consulting firm shows the advertising in the context of ordinary programming (in the case of television commercials) or within editorial material (with print advertising). In some cases, the consulting firm will use concept boards in lieu of finished or rough advertising.

3. Immediately after the advertising exposure, the STM consulting firm administers a short interview or self-administered questionnaire to determine respondent reaction to the advertising. Depending on the simulated test marketing research methodology, the consulting firm will take a variety of measures from respondents.

4. The typical simulated test market research study provides some mechanism for giving the respondent the opportunity to "purchase" the test brand, depending upon the shopping environment's exact set-up. The specific type of "purchase" opportunity depends upon whether the STM uses a simulated or an actual store. In the case where the study employs an in-home usage test rather than a simulated or an actual store, a purchase experience is not available to the respondent. Purchase intent serves as a starting point for trial forecast.

5. Those STMs that use a simulated store or an actual store to allow respondents to purchase the test brand typically employ a purchase incentive. These incentives can take several forms: a coupon for the test brand, a cash payment that can be used to purchase the test product, and a price discount on the test brand and competitive products, among others.

6. For STMs that employ a simulated store, the consulting firm may conduct a post-purchase evaluation to determine the respondents' reactions to the test and competitive brands and to determine why the respondents did or did not buy the test product. The interview may ask a variety of questions, depending upon the
makes are generally more important than the data collection method per se.

Sales Forecasts and Diagnostic Information

Once the STM consulting firm has collected all of the information, the model generates sales forecasts and provides diagnostic information. Some STMs deliver forecasts of consumer awareness, and all provide forecasts of the trial rate, the repeat purchase rate, and market share or sales volume (or both). Each STM uses its own mathematical or statistical model, combined with the use of the STM analyst’s judgment and a database of norms from previous new brand introductions to estimate the new brand’s sales and market share. The consulting firm may provide other forecasts—such as sales forecasts geared to alternative sets of marketing plans or forecasts of market share at the “steady state” or “equilibrium” market—depending upon the specific STM. Diagnostic information may include a variety of tools, depending on the approach the STM uses and the options the client buys. Other information available includes diagnostic reports structured to provide feedback on advertising effectiveness, profiles of the triers and non-triers of the test product, brand attribute analyses, and competitive draw analyses.

In addition to including sales waves as an add-on to the basic simulated test market research design, other STM features are typically available at an additional cost to the client. These additional features include a positioning analysis of the test product and competitive brands (some STMs include perceptual maps), additional reports on information obtained from triers and non-triers, and the evaluation of alternative marketing strategies and tactics.

A Comparison of the Major Simulated Test Market Procedures

While most STMs have characteristics similar to others, each has its own unique “personality,” due to its differences. We have been focusing on the similarities among the major STMs. We now consider the unique features of the five major STMs available to marketers.

In addition to the articles cited earlier, we consulted review articles published by Robinson (1981), Shocker and Hall (1986), and...
Wilson and Smith (1989). Other academic treatments of STMs are also available in textbooks on new product development, product management, and marketing modeling.

Another major source of information is a variety of brochures that STM consulting firms have prepared, published, and distributed. But, despite these sources, the specific details on certain aspects of some of the STMs are still somewhat incomplete. Nevertheless, whenever possible, we will provide enough information on the various STMs to indicate the differences among them.

**Test Product and Packaging Requirements**

As we’ve said, the client must provide the STM consulting firm with the test product, test product advertising, and competitive product advertising. Since all STMs judge consumer responses to a test product, their forecasts can only be as accurate as the test product is representative of actual production. It should be as close as possible to the way it would appear on the shelf in a national rollout. This means that many aspects of the product design, such as the logo, labels, packaging, packaging graphics, and directions for preparation and use, must be finished. Also, it is important that the test product be of production-line quality (especially food products, which often taste better when prepared in small quantities in a test kitchen than when in full production).

ASSESSOR’s first two phases, Concept Evaluation and Screening and Product Positioning, do not require test product. Later phases do require it.

BASES can use packaging prototypes that include the brand name, logo, and directions for preparation and use. The test product must be production-line quality.

LITMUS can use a shelf mockup unless the packaging is critical to the test’s success (as in the case of cosmetics products), when it must be in its most finished form.

Simulator ESP uses the product in its most finished form (as close as possible to the way it would appear on a shelf nationally).

**Advertising Requirements**

The test product advertising applied during simulated test market research should be as close as possible to the actual introductory advertising campaign. The further the advertising is from the finished design, the less accurate the forecasts.

ASSESSOR’s lab store simulation requires that respondents be exposed to test product advertising in a competitive context and therefore requires finished commercials and competitive ads. Several studies have used print-only advertising.

BASES works with a finished concept board that has a photograph of the package or of the product (or both) along with a product description and the key selling messages. It may also use rough or finished commercials as well as animatics.

DESIGNOR works with a concept board, print ad, TV animatic, or final television commercial.

LITMUS uses the methodology that employs print-only advertising in those cases where the client plans a print campaign.

Simulator ESP model prefers finished commercials but frequently uses concept boards.

These five major STMs handle the inclusion of competitive commercials quite differently.

ASSESSOR, DESIGNOR, and LITMUS require competitive commercials if the client plans to use television advertising. The form of the competitive clutter depends on the form used in the test product advertising.

BASES does not use competitive advertising at all.

LITMUS and Simulator ESP also run commercials for non-related products as clutter commercials.

**Other Information Required**

While these five STMs share many similarities in the information they need from the marketing plan, there are also some key differences depending upon the specific study design. With regard to brand awareness, for example, ASSESSOR projects total in-market awareness from client plans for marketing support and expenditures. It does not require an awareness estimate from the client.

Similarly, LITMUS, DESIGNOR, and BASES do not require awareness data to be input since the models generate awareness forecasts as part of their output.

Simulator ESP does not deal with awareness at all since it uses actual advertising and promotion spending dollars (media equal-
ized to reflect effectiveness of GRPs, daypart, and timing) input into the model. There is no need to make an estimate (trial) based upon the estimate of another intermediate variable (awareness).

Each of the five requires estimates of the advertising budget's total dollars and gross rating points. BASES and LITMUS also require a copy of the media plan, and LITMUS requires estimates of total advertising dollars for both the test product and its competitors to generate share-of-voice inputs.

For promotional spending, ASSESSOR, BASES, and LITMUS require estimates of the total dollars to be spent, as well as the number and kind of consumer events to be used. In addition to these data, BASES and LITMUS require a schedule of the timing of these events. Simulator ESP requires estimates of total spending as well as specific details on the product sampling plan to be used (if applicable), such as the percentage of households sampled and the definition of the sampled population.

For distribution build data, all five of the STMs require estimates of all-commodity volume (ACV). Simulator ESP requires "steady state" estimates, while the other models require estimates of ACV in relevant channels at four-week intervals (BASES) or at one-month intervals (LITMUS). ASSESSOR models both distribution and awareness on a period-by-period basis throughout the first year.

LITMUS requires a measure of the total size of the product category, usually generated from Nielsen store audit data. ASSESSOR says that it is helpful to have a client's estimate of total category volume, but it uses the data primarily to check the correctness of the model parameters. (The original ASSESSOR was a share-based model, but this is no longer true. ASSESSOR directly estimates product volume.) BASES and Simulator ESP do not require any estimate of the size of the market for the product category.

Design of the "Shopping" Environment

The shopping experience can be structured in many different ways to simulate a real shopping trip. In fact, each of the five stimulated test market research methodologies approaches the design of its "laboratory" differently.

ASSESSOR generally uses mall-intercept; sometimes it pre-recruits. It may or may not use a simulated store laboratory.

DESIGNOR uses a simulated shopping shelf (mock-up store) or photos showing each product with its price. The firm gives customers a voucher they can use to buy a product or redeem for cash. Any difference between the product's purchase price and the value of the voucher is either refunded or made up by the customer.

LITMUS also uses a simulated shopping environment with all research conducted in central locations in appropriate markets.

The simulated store environments used by ASSESSOR and LITMUS, whether located in a traveling laboratory or in a central location facility, are designed as mock, mini-stores having grocery-store shelving, refrigerators, or frozen-food cases. The mini-store is designed to resemble a supermarket or drug store in which the STM consulting firm can control shelf facings, shelf position, and price. In this way, the company can simulate either expected in-store situations or alternative marketing strategies. With central location facilities, both ASSESSOR and LITMUS offer clients multiple cities in which to collect data. They can use these to gauge reaction to the new brand in various parts of the United States (as well as internationally).

BASES uses no shopping experience at all. Rather, the approach uses consumer reactions in the form of purchase intent, purchase frequency, and the like as the starting point in the sales estimate. The researcher exposes respondents to the test product (or a test concept) in interview facilities in large shopping malls in major markets. After the consulting firm interviews respondents, it gives those who express purchase interest the test product to try at home.

Simulator ESP focuses on an actual in-store shopping experience in a real store (such as Kroger's or Walgreen's). It conducts the testing over a three- or four-day period. Trial purchase of the test product is typically in at least three supermarkets in three different cities. In tests for products such as health and beauty aids or over-the-counter drugs, Simulator ESP may add drug stores to the supermarket outlets. In each store used in the test, the consulting firm sets up a temporary research facility near the store entrance to expose respondents to the advertising and to conduct interviews.
Recruiting Procedures

All five STM methodologies collect a number of demographic characteristics along with product usage information and security measures. Typically, the sample group's final characteristics match as closely as possible the characteristics of the product category's users or the overall U.S. population. To assure adequate coverage of all demographic groups, STM consulting firms commonly resort to quota sampling. Samples may include primary adult buyers (both men and women), other adult buyers, teenagers, or children (or both), depending upon the test product's market characteristics.

ASSESSOR and BASES employ mall intercepts to screen respondents using a personal interview.

LITMUS methodology recruits product category users or brand decision makers over the telephone.

Simulator ESP also uses personal interviews, but the interview is conducted using a store-intercept rather than a mall-intercept technique.

Typical Sample Sizes

The size of the typical sample used in simulated test market methodologies varies widely, depending upon the specific STM.

ASSESSOR sample size is a function of the number of "cells" in a study—variations in the product or its positioning, for example. The sample is larger if a study requires multiple callbacks since respondents drop out at each wave. The company generally wants about 300 respondents per cell and it expects to reconnect 80 percent for follow-up.

BASES—an initial interview with 400 or more respondents, with callback data collected from 190 triers.

DESIGNOR—250 to 400 respondents

LITMUS—400 to 500 respondents initially.

Simulator ESP—approximately 1,500 to 2,000 respondents interviewed in the store with at least 300 triers included.

The Initial Interview

ASSESSOR interviews are person-to-person, using computer entry. The exceptions are self-administered questionnaires following advertising exposure in simulated-store studies. The company conducts most studies in person; it has used telephone (e.g., phone-mail-phone) interviews for selected situations.

BASES conducts a brief initial interview as part of the mall-intercept procedure. Questions pertain to category and brand usage for several product categories, including the test brand category.

DESIGNOR conducts an initial interview in a central location among target consumers for the new product (usually users of its category of products).

LITMUS respondents complete a self-administered questionnaire regarding their current product category and brand usage. Questions also include information on attitudes and preferences about various brands in the category. Attitude measures include specific important attributes for both the test brand and the leading competitive brands. The consulting firm later analyzes these attribute measures to provide guidance on product positioning issues.

Simulator ESP takes a similar approach to that of BASES. During the store intercept interview, the consulting firm questions each respondent briefly concerning brand and category usage for the test product as well as several other products.

Exposure to Advertising

ASSESSOR respondents, after the initial interview, view advertisements for the test product and for the leading competitive brands in the product category. Respondents view these advertisements individually. If the study uses television commercials—now the exception—the commercial for the test product is embedded among four or five competitive spots. If the study uses print advertising, the consulting firm gives a brochure containing the test advertisement plus four to five competitive ads to each respondent.

BASES respondents are not generally exposed to advertising. They are exposed to a concept board describing the new product with no competitive context.

DESIGNOR respondents are exposed to advertising for the new product among advertising for competitive products in the category. The method depends on the quality of the communication available for the new brand. A lower-quality TV ad (e.g., animatics), for example, is shown in isolation, followed by exposure of a reel of
competitive TV ads. A high-quality print ad is inserted in a folder of print ads.

LITMUS respondents view the commercials in a theater-like setting in the context of a television program (usually a situation comedy, with normal commercial breaks). Commercials are shown for the test brand, for the leading brands in the product category, and for non-related products (i.e., clutter commercials) as well. If the client plans to use print advertisements, the STM consulting firm gives a portfolio containing the test ad and four or five competitive ads to each respondent. If rough or finished ads are not yet available, respondents might see a concept board.

Simulator ESP respondents typically view a series of television commercials without other programming. Alternatively, respondents see a concept board for the test product. These commercials or print ads are viewed independently.

Reaction to the Advertising or Concept

Immediately following exposure to the advertising, the STM consulting firms interview respondents to determine their reaction to it.

ASSESSOR uses a short, self-administered questionnaire to measure respondent reactions. The questionnaire includes measures of brand recall as well as recall of the advertising's major messages.

BASES uses a short personal interview to determine reactions to the concept and the respondents' intent to purchase.

DESIGNOR uses a short in-person interview to measure advertising impact, differentiation, and quality. The questionnaire includes measures of brand recall, advertising themes recalled, uniqueness, and relevance scores.

LITMUS uses a short questionnaire that focuses on measures of advertising intrusiveness.

Simulator ESP includes no post-advertising questioning except in the case a Communication Evaluation option is added to the design (available at an additional cost to the client, and depending upon the redemption rate of the coupons provided to the group of respondents). In this case, the consulting firm surveys 100 coupon redeemers and 100 non-redeemers by telephone regarding advertising recall, reaction to the advertising, product category and brand usage, and demographics.

The Shopping Experience

After exposing respondents to the advertising and administering a brief post-advertising questionnaire, the consulting firms give respondents the opportunity to shop in the "store" environment used by the particular STM.

ASSESSOR respondents shop in the simulated store using coupons of sufficient value to permit respondents to make any reasonable purchase.

BASES does not use a store shopping simulation per se. Instead, it uses consumer buying intentions to model trial.

DESIGNOR respondents shop in the simulated store using coupons of sufficient value to permit any reasonable purchase.

LITMUS respondents shop in a store designed to simulate a real-world shopping occasion. They shop as part of a small group and complete an order form for their purchases, a procedure designed to ensure that respondents are making a purchase decision independently of other respondents. The purchase is made at a regular cash register, and the respondents use their own money to pay for it.

Simulator ESP uses real stores, and respondents conduct their shopping "as usual" in their regular supermarket. The STM consulting firm places the test product on the shelf in the appropriate section of the store with other brands in the product category.

The Purchase Incentive

Those STMs that include a "shopping" experience also offer a purchase incentive to encourage respondents to buy the test or a competitive product in the store. The purchase incentive does not apply in the BASES approach since respondents have no opportunity to shop in a BASES study.

ASSESSOR respondents use coupons to purchase a product in the simulated store.

LITMUS respondents are given a discount—generally 30 percent—on all products purchased. While respondents may choose to buy the test product or competitive products or both, they are also told that they do not have to purchase any product (thus simulating a real-world buying situation).
Simulator ESP gives respondents a coupon worth 20 percent off the test brand. For other brands in the category, there generally seems to be no purchase incentive.

**Post-purchase Evaluation**

ASSESSOR, DESIGNOR, and LITMUS use a post-purchase evaluation to question the respondents about their reactions to the test and competitive brands. Also, the post-purchase evaluation attempts to determine why respondents bought the test product or why they did not buy it. This post-purchase interview takes place immediately following the shopping experience.

ASSESSOR interviews roughly 10 percent of the respondents who bought a brand in the product category other than the test brand (i.e., non-triers of the test brand) individually to determine why they bought a brand other than the test product. Also, if the client chooses to purchase perceptual mapping as part of the ASSESSOR study, the consulting firm asks all respondents to rate the brands in the category on a set of attributes.

LITMUS has approximately 5 to 10 percent of the total sample (both purchasers and non-purchasers of the test brand) participate in post-purchase focus groups. Here, the discussion centers on the respondent’s “in-store” experience. To gain insight about the shopping experience, questions probe why respondents did or did not purchase the test brand as well as their reactions to the advertising, the packaging, the price, and so forth. All other respondents (that is, the remaining 90 to 95 percent of the total sample who do not participate in the focus groups) complete a self-administered questionnaire that focuses on the advertising, the packaging, and the price of the test brand and competing brands. Also, this questionnaire assesses pre-usage product positioning by asking respondents to rate the test brand vis-à-vis competitive brands on critical attributes.

Simulator ESP conducts no post-purchase questioning in the store.

**Home Usage Experience**

Since respondents who buy the test brand in the store are considered to be “triers” of the new brand, the STM consulting firms typically ask them to use the brand at home as they normally would use any brand in the product category.

ASSESSOR provides respondents who do not buy the test product with a sample of it so that they can use it at home.

BASES identifies only those who indicate a positive purchase intent in response to the new product concept as triers and gives them the test brand for in-home use.

LITMUS also identifies test brand purchasers as triers. The triers take the product home as they normally would after a normal shopping experience. To avoid bias, LITMUS researchers do not inform them that it will be calling about the product. As an additional feature of the LITMUS approach (available as an option at an additional cost), some of the respondents who do not buy the product may be given the test product for trial at home. This simulates the distribution of samples for the new product. This additional cell is then used to provide an estimate of the impact of sampling on trial sales. In other words, a multi-cell study can forecast the test brand’s incremental sales due to the sampling program.

Simulator ESP study identifies those respondents who redeem their 20-percent-off coupons for the test brand as triers. Simulator ESP will add as an option “Non-trier Diagnostics” to the design; this asks a sample of non-triers to use the test product in their homes.

**The Follow-up Interview**

After enough time has elapsed to allow the respondent to try the test product in a normal usage situation, a telephone follow-up interview determines reactions to the test product to estimate repeat purchase probabilities. Also, the STM consulting firm may also collect information regarding specific product attributes during this interview.

ASSESSOR conducts this phone interview long enough after the simulated shopping experience to allow product usage. This period varies by the nature of the product and can be as little as three or four days. The consulting firm asks respondents for their general reactions to the test product and their after-use acceptance. Respondents also provide their attitudes regarding the test product as well as their preference ratings for the test product relative to a set of competitive products. ASSESSOR measures repeat purchase probabilities...
as well. As an option available at an additional cost, ASSESSOR will arrange focus groups among different groups of respondents to obtain qualitative feedback on certain issues relevant to the client.

BASES surveys triers by phone within an appropriate usage period (two to four weeks) regarding future buying intentions and intended frequency of purchase. BASES also obtains data on price/value considerations and on product satisfaction evaluations. The interview may also include other questions, dealing with respondents' reaction to the test brand on specific product attributes.

DESIGNOR follows up by telephone a few days or weeks after the customer's shopping experience to check product usage, repeat purchase opportunities, purchase intent, purchase frequency, comparison to usual brand/expectations, preference of test product after use versus evoked set, PERCEPTOR attribute ratings of product after use, and other questions.

LITMUS uses a phone study one to six weeks (depending on the product category's purchase cycle) after the simulated shopping experience. The questionnaire is designed to ascertain respondents' reactions to the test product, their repeat purchase probability, their anticipated quantity purchases and usage frequency, and their comparisons of the test brand with other competing brands. In addition, the study collects respondents' ratings on critical product attributes.

Simulator ESP uses a follow-up phone interview (conducted several weeks after the in-store experience), with a base sample of triers. The firm questions triers about their test brand usage, likes and dislikes regarding the test brand, the likelihood of their buying the test brand as a substitute or a replacement for their current brand(s) in the product category, and ratings for several specific product attributes. If the "Non-trier Diagnostics" option is being used as part of the study, the firm surveys a sample of 100 non-triers by telephone within twenty-four hours of the in-store shopping experience. Questions attempt to ascertain advertising recall and reactions, brand and product category usage, reasons for non-purchase of the test brand, and basic demographics.

The Sales Waves Phase
When the STM consulting firm conducts sales waves as part of the follow-up interview, respondents have the opportunity to buy additional units of the test brand at the full retail price. Respondents spend their own money for the product, and multiple sales waves can measure likely depth-of-repeat for the test product. A positive response provides a measure of continued interest in the brand.

STM brochures discuss sales waves, but only briefly. Each of the five STM approaches considered here uses sales waves as an option that can be included in the study design at an additional charge. Additional information is available in the material made available by ASSESSOR, Simulator ESP, and LITMUS.

ASSESSOR includes multiple sales waves in the design when extended use of the test product is an important issue faced by the client. An example would be the client who is concerned the users may lose interest in the test product after the first few usage occasions.

LITMUS usually recommends at least one sales wave for food products and innovative or novelty products.

Simulator ESP has three or more sales waves in 10 to 20 percent of its new product studies. These are used primarily for innovative or novelty products where repeat purchase or frequency of purchase (or both) are hard to predict. In such cases, a series of sales waves can be used to determine the purchase cycle length or to identify potential depth-of-repeat problems.

Sales Forecasts and Diagnostic Information
Once the stimulated test market research consulting firm has obtained all the input data from the respondents participating in a STM study, the consulting firm generates sales forecasts and other variables and prepares diagnostic information. Each model generates parameter estimates using a combination of methods. These include:

1. A mathematical or statistical model designed to provide unbiased estimates of the quantitative relationships that serve as the basic structure of the model.
2. The judgment and experience of the STM analyst.
3. A database of norms from previous new brand introductions (which is often sorted and used on a category-by-category basis).
While it is clear that these three all serve as important contributors to the process of developing each STM's forecasts and diagnostics, the mathematical or statistical model is generally the most important contributor.

As we mentioned above, the academic literature has disclosed much of the model specification and measurement procedures for the ASSESSOR, Simulator ESP, and LITMUS models. These details have also been published for NEWS/Planner, another solid, but less widely used, simulated test market model. However, such disclosure for BASES does not exist.

Sales (and Other) Forecasts

Brochures and other information sources for the five major simulated test marketing research models provide insight into the array of forecast measures that they provide.

ASSESSOR makes the following available:

• First-year sales volume or market share potential or both.
• Projected selling rate at the end of the first year.
• On-going sales potential after the first year.
• Projections for alternative marketing plans and programs.
• Projected sales impact of alternative competitive response scenarios.

The model also projects sources of business for the new product/service, line extension, or restaging, including:

• Cannibalization of the company’s existing business—i.e., contributions from the company’s existing products or services.
• Projection of incremental business—i.e., contributions from sources other than the company’s existing products or services.
• Projection of business contributions from demographic groups, behavioral segments, attitudinal segments, and retail trade segments.

BASES output includes:

• Sales volume in the first year (with an option for the second year).
• The trial rate (for the first year).
• The repeat rate (for the first repeat purchase).
• Repeats per repeater.
• The average time between purchases.

The average number of units per trial and repeat purchase occasions.

Awareness.

DESIGNOR output includes:

• The trial rate.
• The adoption rate (or retention).
• The share of choice (or share of purchases).
• Year 1 and Year 2 volume sales build-up.
• Long-term stabilized share (Year 2 and beyond)
• Price elasticity
• Cannibalization/draw estimates.
• Awareness modeling. Cross-checking/correcting client assumptions.

LITMUS provides:

• Monthly awareness (for the first year with an option for the second year).
• Monthly market share or sales volume or both (for the first year with an option for the second year).
• The trial rate.
• The repeat purchase rate.

Simulator ESP provides:

• Sales volume (for the first year).
• Monthly trial rate, repeat purchase rate, and sales volume.
• The percentage of the target market who will “ever” repeat the purchase.

Diagnostic Information

The diagnostic information that these simulated test market research models generate include a variety of different items, depending on the STM's approach and the options the client purchases. Accordingly, some of the diagnostics available for each STM are provided below.

ASSESSOR is capable of providing the following:

• An evaluation of advertising or concept communication, covering major selling points: differentiation versus competitive prod-
ucts or services; perceived advantages and disadvantages versus major competitors; and reasons for trial motivation or rejection.

- Product/service fit with the advertising or concept, covering perceived delivery versus expectations; reasons for satisfaction or dissatisfaction.

- Product/service acceptance, covering perceived advantages and disadvantages of the product overall and relative to existing competitive alternatives; importance of product delivery dimensions on acceptance and adoption of the product; preference and intensity of appeal relative to competitive products.

- Consumer adoption diagnostics covering trial potential; first repeat purchase potential; on-going repeat purchase potential; purchase/usage cycle; quantities purchased per occasion; average purchase price.

- Impact of the marketing plan and program in creating awareness; generating trial/first purchase; and developing repeat purchase and adoption.

BASSES diagnostic information includes:

- Respondent likes and dislikes of the concept.

- Profiles of triers, focusing on demographics, product categories and brands used, and brands on hand.

- Profiles of the test brand, centering on respondents' likes and dislikes, areas for improvement, attribute ratings, and product usage and directional ratings.

DESIGNOR diagnostic output includes:

- Measures of the brand's IDQV—impact, differentiation, quality, and value. Impact is the brand's capacity to create strong identification with the target consumer via advertising and packaging. Differentiation is the brand's capacity to create a dominating position of both uniqueness and some perceived consumer benefit via the brand's mix. Quality is measured by the product's pre-use perception, importance, and usefulness to the consumer; the anticipation of the quality performance; the positive expectations generated by the product's positioning; and the post-use performance when the customer judges the product against the promise. And value is the relationship of total quality and price in the competitive environment.

LITMUS's diagnostic output includes:

- Advertising effectiveness, including brand recall from the post-advertising (but pre-shopping) period, brand recall and main message communication measures from the post-purchase questionnaire, and brand recall and main message communication measures from the focus group discussion.

- Target market guidance, including demographics, category and brand usage, and measures of brand preference measures for the test brand and leading competitors.

- Product guidance, including respondents' usage, likes and dislikes, repeat purchase intentions, and reactions to the test brand's name, label, package, and price.

- Marketing plan diagnostics, including a sensitivity analysis to changes in the various marketing plan input levels, a "source of volume" analysis, and marketing mix optimization plan.

Simulator ESP's diagnostics include:

- Measures of advertising effectiveness (available as an option).

- Profiles of triers, including demographics, categories and brands used, and a definition of the "interested universe" or target market.

- Profiles of non-triers (available as an option).

- Profiles of the test brand, including respondent usage patterns, likes and dislikes, repeat purchase intentions, post-usage attribute ratings, and likely replacement or substitution for the current brand(s) purchased.

- A report on the introductory spending level necessary to achieve sales objectives.

Additional Features that Are Available

In addition to including sales waves as an add-on to the basic simulated test market research design, other STM features are available at an additional cost to the client.

ASSESSOR and BASSES clients, for example, can obtain a forced-trial report, which focuses on ways to encourage concept rejectors to try the test brand (for example, through couponing or sampling).
LITMUS and DESIGNOR include any of the following at an extra cost: cannibalization analysis, competitive response analysis, price sensitivity, and market share optimization.

Simulator ESP can make available a report on non-trier diagnostics (using a sample of 100 respondents). Also, a Communication Program report is available, which estimates the impact of a free sample (instead of the coupon that a Simulator ESP study typically uses) on the trial and repeat purchase rates.

Validity

Each of these five simulated test marketing research methodologies provides an indication of the predictive validity of its studies. The claimed accuracy for each of the models is as follows:

ASSESSOR forecasts in Phases III, IV, and V have been within plus or minus 10 percent in approximately 90 percent of the cases where products/services were evaluated with the consumer having the opportunity to experience the product or service. In earlier stages of development where consumers respond only to a concept or advertisement, accuracy has been plus or minus 20 percent approximately 80 percent of the time.

BASES reports that in-market sales have been within an average of 10 percent of predicted sales in 250 validated cases.

DESIGNOR says that in-market sales have been on average less than 9 percent of forecast sales. In the United States, almost two-thirds have been within 10 percent deviation.

LITMUS says that in-market sales have been within 10 percent of predicted sales in nine out of ten of the approximately 200 cases that have been validated.

Simulator ESP says that for forty-eight verified cases, in-market sales have been within 10.5 percent of the model’s forecasts.

Timing and Cost Estimates

These five consulting firms indicate the following timing—from field start to final presentation—and cost estimates at the end of 1993:

ASSESSOR: Phase I and II, six to eight weeks; Phase III, eight to ten weeks; Phase IV, nine to twelve weeks; and Phase V, three to five months. Costs vary with the phase, the number of cells in the study, the effective incidence for qualified respondents, and whether or not extended use is incorporated. Costs can range from $45,000 up.

BASES: approximately $50,000 to $60,000. The consulting firm provides a volume estimate eight weeks from the field start, and a final report ten to twelve weeks from the field start.

DESIGNOR: approximately $80,000 for a concept test to $150,000 for a full DESIGNOR test. Typically a study takes eleven weeks from commissioning to presentation.

LITMUS: approximately $175,000, plus $15,000 per sales wave. Timing is twelve weeks from the field date.

Simulator ESP: approximately $50,000 to $60,000, plus $12,000 for three sales waves, plus $5,000 for each set of non-trier diagnostics, plus about $500 for each additional model run to address the appropriate level of introductory spending. Timing is twelve to fourteen weeks from the field date. The consulting firm provides top-line forecasts approximately six weeks from the field date.

Management Implications

In systematically comparing the five major STMs, several conclusions can be drawn. Some of the more salient are the following:

ASSESSOR (and, by extension, DESIGNOR), Simulator ESP, and LITMUS have all been documented in the academic literature, where the focus of several articles on each has centered on issues such as the structure of each model and measurement issues. Unfortunately, this type of information is not available for BASES; virtually none of its details have been published for industry scrutiny.

There are major differences among the five methodologies regarding how the consumer’s in-store experience is “simulated.” These differences may be important in obtaining high-quality results from the STM study.

The details of the collection of data from respondents vary widely, depending upon the STM. Marketing managers must consider these differences carefully before they select an STM procedure since some introduce a large amount of artificiality into the study and thus may limit the results and conclusions that may be drawn from them.
Simulator ESP does not deal with brand awareness data. On the other hand, BASES, DESIGNOR, and LITMUS generate brand awareness forecasts and consider these forecasts to be an important outcome of the modeling process.

ASSESSOR and Simulator ESP provide forecasts for a “steady state” or “equilibrium” period. LITMUS and BASES, however, provide forecasts for Year One (and sometimes Year Two) quantities and are dynamic models of market event timing. This approach more clearly allows managers to assess the incremental effects of advertising, sales promotion, and distribution on consumer awareness, trial, repeat purchasing, sales, and market share.

The ASSESSOR model seems to be primarily a “market share” model. This means that, in many cases where the product category is ill-defined, ASSESSOR has difficulty in predicting unit or dollar sales.

A major difference between the systems is in terms of their diagnostic capabilities. DESIGNOR, as an illustration, employs interesting technology for identifying the perceptual dimensions that determine brand preference; Simulator ESP can readily assess the effects of in-store promotion better than alternative systems; LITMUS enjoys a “sensitivity analysis” and optimization technology that are unique.

The remarkable thing about the five STMs described here is that despite their differences they appear to perform similarly—assuming that we accept the claims of the research companies—in terms of predictive validity. This is not unlike a situation in medicine today where different types of treatment for, say, back pain (treatments ranging from drugs to chiropractic therapies to surgery) appear to yield comparable levels of success.
The Most Important Additional Reading of All

Once you decide to pursue a scientific career, it is essential that you monitor the journals in your field regularly. Most disciplines have a general research journal associated with their primary professional organization, as well as several more specialized vehicles that will routinely publish studies in your general area of interest. If you can afford it, subscribe to a couple of these, but in any case read them routinely as they come out. Also, remain literate with respect to the major scientific policy issues in general by keeping an eye out for articles in the newspaper as well as monitoring such periodicals as The Chronicle of Higher Education and Science.

Index

Abstracting services, 23
Accuracy checking, 128, 137
Alpha level, 131-132
Alternative explanations, 81-91
Analysis of covariance (ANCOVA), 102, 105, 127, 128, 129, 131, 132
Analysis of variance (ANOVA), 105, 127, 128, 129, 131, 132
Aptitude-by-treatment interactions, 133
Artifacts. See Experimental artifacts
Assumptions, 132-133
Attrition, 85-87

Between-subjects grouping variables, 101, 130
Bibliography, 27-28. See also Literature review
Black box approach, 97
Blocking variable, 61, 101, 102, 129, 130
BMDP, 126
Brooks, Mel, 41

Campbell, Donald, 65
Career counseling, 11-12

Causal inference, 47, 62-63
Causal links, 100
Clinical trials, 6
Coding, 127
Combined treatment groups, 97-98
Comparison group. See Control and comparison groups
Compensation, 56
Computerized literature searches, 23
Computer software: statistical packages, 19, 126-127
wordprocessing, 27-28
Confounding variables, 75-77
avoidance/minimization of, 77-80
Control and comparison groups, 59-60
cross-contamination, 90
designing for maximum effect sizes, 95-96, 99
differential subject attrition, 85-87
nontreatment procedural effects, 76
practical problems, 72-74
sample size, 106
Controlling variables, 42, 68, 101-103, 129. See also Covariates
Correlational research, 8, 62-64
Counterbalanced model, 69-70
CONDUCTING MEANINGFUL EXPERIMENTS

Counterbalancing, 78-80
Covariates, 61-62, 101-103, 115, 129-130. See also Controlling variables
Criticism, 52-55
Cross-contamination of subjects, 90-91
Crossover designs, 68-69, 71, 102, 129
Curiosity, 17

Data analysis:
assumptions, 132-133
procedures, 128-134
secondary and exploratory analyses, 133
statistical packages, 19, 126-127
Databases, 23-24
Data entry, 127-128
Dedication, 13-14, 16
Degrees of freedom, 132
Delegation, 61, 122-123
Dependent variables, causal links to, 62-63, 100. See also Outcome variables
Descriptive research, 8
Design, experimental. See Experimental design
Diary, 83, 135
Directional hypothesis, 41
Discussion section, 135
Dissertations, 25
Doctoral dissertations, 25
Double-blind, placebo design, 76

See also Experimental artifacts
Experimental artifacts:
cross-contamination of subjects, 90-91
history effects, 83
obtrusive procedures, 87-89
procedural confounds, 75-77, 81
regression toward the mean, 84-85
selection effects, 82-83
subject attrition, 85-87
testing effects, 89-90
Experimental design, 58-62
confounding (procedural) variables and, 75-80
correlational (nonexperimental) designs, 62-64
counterbalancing, 80-81
effect size maximization, 95-99
intact groups, 67
nesting design, 80-81
quasi-experimental studies, 64-72
sensitivity, 93-94
written description, 121
See also Control and comparison groups; Random assignment
Experimental interval, 96
Experimental studies, 6-8. See also Intervention
Experiment implementation. See Implementation
Exploratory analysis, 133

Factorial analysis of variance (ANOVA), 105, 127, 132-133
False negative results, 60, 81, 93
acceptable levels, 104
procedural confounds and, 77
selection effects, 82
False positive results, 60, 81, 93
acceptable levels, 104
procedural confounds and, 77
selection effects, 82
Feasibility study, 115-117
Follow-up hypotheses, 51
F ratio, 132

Garbage in-garbage out, 127

Index

Good subject phenomenon, 87
Hawthorne effect, 87
History, 83
Honesty, 12, 135-136
Hypothesis, 38-45
follow-up, 51
statement of, 42-45
subjects in, 43-45
testable, 42, 47
types of, 41
Hypothesis testing, 47-55
outcome significance and implications, 49-50
professional criticism, 52-55
research limitations, 48-49
statistical procedures, 105
strong inference, 56-57

Implementation:
delegation, 122-123
evaluation of, 123-125
monitoring, 120-122
Independent variable:
defined, 33
nonmanipulable (correlational studies), 62-64
potential nonintervention factors, 130
selection of, 32-34
See also Intervention
Inference, 56
Informed consent, 88
Intact groups, 67, 82
Interrater reliability, 121-122
Interval design, 96
Intervention, 6-8
designing for maximum effect sizes, 95-99
development of, 109-114
evaluating, 113
independent variable selection, 32-34
preliminary efficacy study, 63-64
See also Independent variable

Jenkins, Joseph R., 5

Journals, 23-24, 136
Learning curve, 89
Literature review, 14, 21-27, 135
annotated bibliography, 27-28
computerized searches, 23
meta-analysis, 29
review article, 28-29

Mathematical skills, 19
Maturation, 84
Meaningfulness, 1-2, 6
of hypothesis, 45, 47-55
outcome variables, 30-32
subjectivity and, 55-56
Mentors, 5
Meta-analysis, 29, 105, 134
Methods section, 135
Monitoring, 61-62, 120-122
Moody, William B., 5
Motivations, 39-40, 50, 111
Multiple regression, 131
Multiple univariate analysis, 131
Multivariate analysis of variance, 131

Naturalistic research, 8
Nesting design, 80-81
Nesting variables, 130
Nonrandom assignment, 66-68
Null hypothesis, 41, 93
Null results, 23

Observational research, 8
Open-mindedness, 18-19
Operational statement, 39, 40
Outcome variables, 6-8
alternative explanations for, 81-91
causal linkages, 62-63, 100
data analysis, 130-131
effect size maximization, 95-99, 106-107, 112
in hypothesis statement, 42-43
learning effects, 69
conducting meaningful experiments

obtrusiveness of, 89
refinement of, 114-115
regression toward the mean, 84-85
reliability, 99-100, 114-115
selection of, 30-32
sensitivity, 100, 115
visualizing, 49-50

Patience, 18
Personal motivations, 50, 111
Ph.D., 11-12
Pilot study, 109-110, 117-119
development of intervention, 109-114
feasibility study, 115-117
outcome measure refinement, 114-115
sample size for, 118
Placebo, 76, 87
Platt, J., 56
Post-intervention survey, 88, 91, 123-124
Practical experience, 15-16
Pre-experimental design, 65. See also
Quasi-experimental studies
Preliminary studies, 7, 63-64. See also
Pilot study
Pretest effects, 89-90, 130
Procedural confounds, 75-77
Professional feedback, 52-55
Proofreading, 128, 137
Publishing, 51, 134-138

Quasi-experimental studies, 64-72

Random assignment, 59-62, 70-71
hypothesis statement, 43
in crossover designs, 69
practical issues, 72-74
procedural components, 77-78
supervision of, 61-62
Randomized block design, 102, 133
Record-keeping, 83, 122
Reference lists, 25
Regression toward the mean, 84-85
Rejection, 138
Reliability, of outcome measures, 99,
114-115

Repeated measures design, 102
Report, 134-138
Research diary, 83, 135
Researcher qualifications and characteristics, 16-19
academic experience, 11-12, 14
honesty, 12, 135-136
practical experience, 15-16
Research hypothesis, 41
Results:
data analysis, 126-134
publishing, 134-138
Review. See Literature review
Review article, 28-29
Rewards, 56

Sample size, 104-106
ccontrol versus intervention, 106
for pilot study, 118
quasi-experimental designs, 67
subject attrition and, 85-87
SAS, 126
Scientists, 3. See also Researcher
qualifications and characteristics
Scripts, 121
Secondary analysis, 133
Sensitivity, 93-94, 115
Single-group studies, 64-66
Skepticism, 17-18
Social Science Citation Index, 26
Software, 27-28, 126-127
Solomon four-group design, 90
"So what?" question, 50, 51
Sphericity assumption, 133
SPSSx, 126
Stanley, Julian, 65
Statistical procedures:
assumptions, 132-133
choosing, 105, 128-134
software packages, 19, 126-127
Statistical significance, 94-95
effect size maximization, 95-99, 106-107, 112
"fishing expeditions" for, 133
power analysis, 104-106
sample size and, 104-106

using pre-existing (controlling) variables, 101-103
Strong inference, 56
Study log, 121
Subject(s):
attrition, 85-87
benefits from intervention, 98
cross-contamination of, 90-91
description in hypothesis, 43-45
exclusion criteria, 98
experimental design, 58-62
maturation effects, 84
obtrusive procedural effects, 87-89
protecting, 16
See also Control and comparison
groups; Random assignment
Subjectivity, 55
Supervision, 61-62, 120-122

Testable hypothesis, 42, 47
Testing effects, 89-90
Time series, 65-66, 70, 72
Training and experience, 11-12, 14, 15-
16, 121-122
Type I error probability, 104
Type II error probability, 104

Unexplained error variance, 101

Within-group variation, 101-103
Within-subjects analysis of variance, 105
Within-subjects factors, 129, 129
WordPerfect, 27-28
Wordprocessing software, 27-28