23

AD TESTING

ALLAN L. BALDINGER
Consultant

WILLIAM A. COOK
Advertising Research Foundation

HISTORICAL CONTEXT

Growing in Use and in Controversy

Advertising research grew dramatically in the 1950s, as television burst onto the media scene. However, in the 1980s, marketers shifted some of their attention towards measuring sales promotions. This was at least partly due to the emergence of weekly scanner-based data on sales. Scanner-based data provided direct evidence of the short-term effects which in-store promotions have on packaged goods brand sales; consequently, spending on ad testing declined, at least in proportion to sales promotion measurement. Whether high times or low, ad testing has been frequently embroiled in controversy.

David Ogilvy and Leo Burnett were great ad men and great advocates of advertising testing. In 1994, David Ogilvy puzzled, "Most creative people today detest research, and I've never understood why. . . . In my day, I used research very often to give me the courage to run campaigns that were risky. My famous eye-patch ads for Hathaway shirts, for example, would not have been created had I not been studying a chart that I saw in Harold Rudolph's Attention and Interest Factors in Advertising."

Today, many creatives would still differ with Ogilvy and take the position of Kevin Roberts, CEO, Saatchi & Saatchi Worldwide, who counters, "Two of the biggest mistakes we make in advertising research involve copy testing and the timing of the research. Despite what our creatives will tell you, not all research kills creativity, but quantitative copy testing certainly does. It does so by attempting to quantify the unquantifiable." Roberts urges creatives to avoid the "overuse of research as a form of judgment on creative work, and its under use as a source of insight into the mind and mood of the consumer" (Roberts, 1998).

The creatives' hostility toward advertising testing grew particularly heated in the late 1970s, when marketers pushed for simplification and a single measure on which to make copy selection decisions. The "one-number-fits-all" mentality was a sharp stick in the eye of creatives seeking to tailor their ads to particular groups of consumers and competitive situations.

Charles Young (2004) has observed that

> there are four general themes woven into the last half-century of copytesting. The first is the quest for a valid single-number statistic to capture the overall performance of the advertising creative. These are the various "report card" measures used to filter commercial executions and help management make the go/no go decision about which ads to air. The second theme is the development of diagnostic copytesting, whose main purpose is optimization, providing insights about and understanding of a commercial's performance on the report card measures with the hope of identifying creative opportunities to save and improve executions. The third theme is the development of nonverbal measures in response to the belief of many advertising professionals that much of a commercial's effects—e.g., the emotional impact—may be difficult for respondents to put into words or scale on verbal rating statements and may, in fact, be operating below the level of consciousness. The fourth theme, which is a variation on the previous two, is the development of moment-by-moment measures to describe the internal dynamic structure of the viewer's experience of the commercial, as a diagnostic counterpoint to the various gestalt measures of commercial performance or predicted impact.

The ad testing being conducted today has evolved and grown substantially since World War II. There is a direct correlation between advertising expense and the expenditures on ad testing. Since the bulk of advertising expense is now devoted to television, ad testing research in the United States is predominantly conducted on television ads.

Growth of the Media

Advertising testing followed the development of the media. As media spending changed over time, so did the research associated with it. Prior to World War II, the predominant media in the United States were newspapers, magazines (or print), and radio. But when television exploded on the scene in the 1950s, advertising research and testing exploded along with it.

Table 23.1 lists some publicly available figures on the relative importance of the various media, in the United States, for the first half of 2004. The total spending was almost $68 billion, which would mean that, on an annual basis, ad spending in the United States is roughly $140 billion.

Table 23.1 Media Spending

Media	*$ Spending ($ Billions)*	*% of Total*
Television	29.6	44
Newspapers	14.4	21
Magazines	13.2	20
Radio	5.3	8
Internet	3.7	5
Outdoor/billboards	1.4	2
Totals	67.6	100

Source: TNS Media Intelligence/CMR (Johnson, 2005).

The fastest growing of these media is Internet advertising, which grew at a +26% rate versus the first half of 2003. Still, advertising on the Internet represents only 5% of total ad spending.

There are no publicly available figures on how the ad testing business itself splits by media. However, it is likely that the proportions differ somewhat from the spending figures shown here. For example, most newspaper advertising is really retailer-based sales or promotion-related ads. These are rarely tested with consumers by the advertisers who buy them since they are merely intended to spur a short-term sales boost, not to communicate an enduring or persuasive sales message. Many newspaper ads are sponsored by small local companies, which lack the financial resources to conduct sophisticated ad testing.

In addition, many ad campaigns start out as TV campaigns, where the bulk of campaign money is spent, and then are spun off into other media as magazines or radio ads, using the same themes and content as were used in the TV campaign. Consequently, many advertisers test only the TV campaign for its likely effectiveness and trust that its effectiveness will be easily translatable into other media. This assumption of equivalent effectiveness in other media as in TV has been found wrong in numerous instances.

Still, many campaigns are specifically oriented toward, and tested within, the nontelevision media outlets, such as radio, print, or the Internet, as will be discussed below.

Evolving Research Philosophy

Although a change appears under way today, ad testing remains dominated by a model called the AIDA model, originally published as an explanation of how personal selling works by St. Elmo Lewis, in 1898. The AIDA model stands for the following:

A Attention

I Interest

D Desire

A Action

The AIDA model is part of the hierarchy of effects theory, which holds that the basic and ultimate objective of all advertising is sales. It is a linear theory, meaning that it presumes that consumers must go through a rational and sequential series of steps. The task of advertising begins with developing "attention," or awareness of the brand being advertised. The second stage is to generate "interest," through the communication of a relevant sales message. The third stage is to create "desire," the ability to persuade the consumer that the communication is convincing enough that she or he is actually motivated to buy the brand. And the fourth stage is "action," usually the actual purchase of the brand being advertised.

It should be noted that sales is not always the ultimate goal of advertising. For example, in automotive advertising, the goal is to create sufficient interest to motivate consumers to visit a showroom. The actual sale takes place once the retail salesman takes over and following a test drive. Similarly, public service advertising often has as its goal a simple shift in consumer attitudes (for example, Smokey the Bear motivated consumers to "prevent forest fires" by being careful with matches in national parks, while antismoking campaigns are intended to reduce the sale of cigarettes, not increase them). However, in virtually all cases, advertising has as its goal at least a shift in "desire."

While the outlines of the AIDA hierarchy are widely visible in most approaches to ad testing today, the hierarchy has grown in layers, variations, and complexity. As early as 1974, Michael Ray (1974) noted that there are three types of behavior exhibited in consumer behavior: cognition, affection, and conation, or thinking, feeling, and doing. Batra and Ray (1986) found that in different situations, the order of occurrence of those behaviors varied—that feelings could precede, and therefore influence, thinking and doing.

The great volume of brain research over the past decade has pushed the interest in emotion and affective response to the front stage. Gerald Zaltman (2003) went so far as to declare that senior managers need to change the way they think about how their customers think:

> The most troubling consequence of the existing paradigm has been the artificial disconnection of mind, body, brain and society. . . . Only by reconnecting the splintered pieces of their thinking about consumers can companies truly grasp and meet consumers' needs more effectively—and thus survive in today's competitive and rapidly shifting business environment.

Eight ad testing companies are participating along with Zaltman, as well as leading advertisers and agencies in a consortium study of the emotional response to advertising by the Advertising Research Foundation–American Association of Advertising Agencies (ARF-AAAA). One of the participants is John Hallward from Ipsos-ASI, a longtime provider and avid defender of recall testing. Hallward (2005) observes, "Since the first impressions of advertising may often be more emotional than rational, we need to explore beyond the rational to better understand consumers' emotions towards and impressions of the product or the message" (p. 3).

MODELS

Models of How Advertising Works

The AIDA model described above has been integrated into numerous models in current use today. They are broadly referred to as *hierarchy of effects* models. As the constructs of consumer attitudes and emotions were added, the models have become more than frameworks for teaching about advertising and have gained increasing relevance to the measurement of advertising.

Bill Wells (1989) observed that ad forms could be described as "lectures" or "dramas" or combinations of the two. A lecture is an ad in which an announcer speaks directly to the viewer through the TV screen. These ads tend to do well on persuasion but less well on attentiveness. Most demonstrations of a brand's specific functional benefits are lectures. On the other hand, an ad that is categorized as a drama is a mini-play. An ad of this type can be highly entertaining and can do very well on measures of attentiveness. However, the creative challenge in these ads is often to ensure that relevant brand benefits are also properly communicated. Wells also introduced an important distinction between the motivational routes of the advertising; rather than simply being informational, ads could be transformational, for example, enhancing the product consumption experience or the esteem one feels when wearing or using the brand (Puto & Wells, 1984).

For the ad agency Foote Cone Belding, an attitudinal model, the FCB grid, was developed that tied together the two constructs, high involvement/low involvement and thinking/feeling, into a simple 2 × 2 grid. Products such as insurance and televisions were placed in the thinking/high-involvement quadrant, while salty snacks and greeting cards went into the feeling/low-involvement quadrant.

Rossiter and Percy (Rossiter, Percy, & Donovan, 1991) expanded on the FCB grid. They added brand awareness in their model and replaced the thinking/feeling distinction with the type of motivation involved in the ad—either informational or transformational. For transformational motives, they included sensory gratification, intellectual stimulation, and social approval.

Media Models

In 2002, the Advertising Research Foundation published a monograph titled *Making Better Media Decisions,* updating the 1961 version, *Toward Better Media Comparisons.* These papers described the various stages in which media are communicated and measured, described below. The purpose of the 2002 update was to reflect the changes in the media environment that had occurred between 1961 and 2002, such as the introduction of the Internet as a new media vehicle and the increased emphasis on sales.

Both models include, as key initial stages, vehicle distribution, vehicle exposure, and advertising exposure. Both models also include sales response as the ultimate and final purpose of advertising. However, the updated model changes advertising perception to advertising attentiveness and adds advertising persuasion and advertising response. In the 2003 model, the first three are descriptions of the media and its distribution, while the last five describe consumers' reactions to the media.

Here are the definitions of each of the eight stages, taken directly from the 2002 monograph, *Making Better Media Decisions* (Advertising Research Association, 2002):

1. *Vehicle Distribution.* This is a count of physical units through which advertising is distributed. It is a pure media effect. Measurement techniques include newspaper and magazine circulation studies, TV- and radio-tuning studies, online media page requests, and billboard locations.

2. *Vehicle Exposure.* This is a count of the people exposed to the media vehicle whose eyes or ears are open. It too is a pure media effect. Measurement techniques include radio and TV people ratings, magazine readership studies, online media page view counts, billboard traffic counts, and so on.

3. *Advertising Exposure.* This is a count of the people exposed to the media vehicle who are also exposed to its advertising. It is the highest level of measurement that is still a mostly pure media effect. Measurement techniques include radio and TV commercial audience ratings, print ad page exposure studies, online ad view counts, billboard traffic counts, and so on.

4. *Advertising Attentiveness.* This is the degree to which those exposed to the advertising are focused on it. It is the first measurement level at which the effects of the medium are significantly confounded with the effects of the creative. Measurements include dedicated attentiveness studies, recall or campaign tracking studies, brainwave research, and so forth.

5. *Advertising Communication.* This is a measure of the information retained by the consumer after exposure to the message. Measurement techniques include advertising and brand awareness tracking, copytesting/recall, advertising recall studies, and so on.

6. *Advertising Persuasion.* This is a measure of the shift in intentions produced by advertising communication. Here we are interested in the medium's ability to frame the message in ways that make it more credible, more relevant, and hence more persuasive. Measurement techniques include advertising tracking, copytesting, intent to purchase, willingness to consider, and so forth.

7. *Advertising Response.* This refers to measures of consumer response short of sales. Examples include visiting a showroom, calling a toll-free number, clicking on an online ad, requesting a brochure, and so on. In direct mail and interactive media, such responses can be measured directly. Measurement techniques include click-through, post-click-through interaction, lead generation, telephone and mail response, coupon redemption, and so on.

8. *Sales Response.* This is purchase of the advertised product or service in response to the advertising. Of all the measures listed, it is the most relevant to the advertiser but the least dependent on advertising and media effects. In addition to sales, useful measures include profits, return on investment (ROI), and consumer lifetime value (CLV)—an estimate of the future profitability of a newly acquired customer. Measurement techniques include sales tracking, test markets, single-source panel research, and marketing-mix modeling.

The ARF media model provides advertisers, the media, ad agencies, and research companies a useful framework to address the research complexities and issues involved in the measurement of advertising effectiveness.

Table 23.2 provides the types of research most relevant to the individual stages. Stages 1 through 3 are addressed by media research measures of media delivery and audience exposure. Stages 4 through 6 are the areas where copytesting techniques are applied, while Stages 7 and 8 are those for which ad tracking and in-market sales tracking methods are used.

Table 23.2 Types of Research

2002 ARF Media Model	*Research Type*
1. Vehicle distribution	Media research
2. Vehicle exposure	Media research
3. Advertising exposure	Media research
4. Advertising attentiveness	Ad/copytesting
5. Advertising communication	Ad/copytesting
6. Advertising persuasion	Ad/copytesting
7. Advertising response	Ad/copytesting/in-market tracking
8. Sales response	In-market tracking

The industry has long desired the availability of techniques that provide a "single-source" method of advertising measurement (e.g., a single system that links media or advertising exposure to sales response).

Advertising Effects Models

In the context of marketing research, a "model" typically refers to an analytic approach to a given data set, designed for the purpose of adding value to the data. Models vary widely in complexity, approach, and scope. They range from purely descriptive models, such as the ARF media model just described, to highly technical and sophisticated models employing advanced statistical techniques.

Sometimes, the same research project or study will provide useful measures of several stages of the ARF media model. However, measuring the effectiveness of a particular advertising campaign is not a simple task. Since a given brand may be spending money on multiple media outlets and advertising campaigns simultaneously, as well as event sponsorships, public relations (PR), trade, and customer promotion, teasing out the actual ROI of an advertising campaign or expense can be very difficult.

To deal with this complexity, many of the research companies involved in the measurement of advertising have developed their own models of advertising effectiveness. These can

vary considerably in their complexity and from company to company. Indeed, sometimes the same research company will employ different modeling approaches depending on whether a given research project is a copy pretest, a survey-based ad tracking project, or an in-market evaluation of long-term spending versus sales trends.

There is considerable published literature covering the validation efforts by various research companies, much of which discusses the general modeling approaches used (e.g., Adams & Blair, 1992; Lukeman, 1995). However, the technical details of the various models often remain within the research company, as a proprietary asset of the company. These technologies are frequently a key to the company's differentiation in the crowded ad testing marketplace.

The Stages of Advertising Research

Developmental Research and the Qualitative Role

Research is performed at the various stages of the advertising process. However, it is estimated that well less than half of the ads in the media each day have been developed through a rigorous process of research.

Research can contribute to advertising at several points of development:

- Copy development stage
- Rough commercial stage
- Final production stage, prior to air
- After airing

The copy development stage is the initial stage in the development of new advertising of a product or service. Often, ad agencies talk about this stage as a copy exploratory. A very common research method used at this stage is focus groups or other forms of qualitative research. Small groups of targeted consumers are recruited, usually by phone, and are brought into central location facilities. This is the stage at which the client's ad agency is most directly involved. In fact, many copy exploratory research projects are paid for by the ad agency.

The main purpose of copy development/exploratory research is to help identify the optimal advertising strategy to employ in the new advertising. The strategy represents the underlying communication objective of the ad. It represents a summary of the claim or claims to be communicated about the brand and the supporting evidence for that claim. The execution is the advertising agency's creative interpretation of how to most effectively translate the selected strategy into an actual advertisement.

A typical strategy statement for an ad campaign might look something like this:

> The (BRAND) is better at (CLAIM/BENEFIT), relative to (COMPETITIVE FRAME), because it (SUPPORTING EVIDENCE).

A completely hypothetical example might be the following:

> Rolaids is better at curing heartburn within 20 minutes, relative to Tums, because it contains a new and improved active ingredient.

In a typical developmental qualitative project, a variety of possible claims and/or supporting evidence will be shown to possible brand buyers and/or current brand buyers. Insights will be obtained about the relative merits of these benefit statements, product claims, and potential supporting evidence. It is usually not advisable to settle in on one final strategy or claim, based on group sessions alone. That is because group sessions, by their very nature, employ small sample sizes of consumers and because a variety of possible biases can creep into such sessions. One type of bias of concern to those using focus group research is leadership bias, which occurs when the session's respondents merely agree with a particular strong or articulate group participant. Another is social acceptability bias, in which case respondents voice opinions based on social norms, rather than their true feelings (e.g., "I never buy any brands based on advertising.").

The general purpose of such qualitative sessions is to reduce the number of possible strategic alternatives to a smaller, more easily testable number of alternatives.

Many research companies have also created quantitative early stage approaches. In these techniques, alternative advertising strategic approaches can be ordered on their predicted effectiveness based on large sample sizes of consumers who are exposed to advertising stimuli in the form of simple statements of claims or as rough print ad–style concepts.

Evaluative Pretesting

Once the copy exploratory developmental phase has been completed, the marketer will usually move in the direction of testing the advertising, using quantitative methods. This is often called copy pretesting, with the *pre* referring to the testing of advertising prior to actually placing the advertising in the media.

Depending on whether the number of advertising executions is large or small, as well as whether the likely costs of production of the advertising are high or low, the client and ad agency may either test the advertising in "rough" or "finished" form. In other words, an advertisement may be inexpensively simulated by means of a "rough" version of the advertisement.

The Value of Norms

Once an advertiser decides which ad testing measures are likely to be the most predictive of marketplace success for his or her brands, normative databases become extremely valuable. The advertiser knows that some of his or her ads have been more successful than others in the past and often believes that ad campaigns must be frequently changed because of "wearout." But he or she also knows that advertising success will vary by product category and even from brand to brand within the same category. Therefore, a database of scores, gathered consistently over time, using the same technique and sample specifications, is the best means of separating better ads from weaker ones (Brandt & Biteau, 2000, p. 25).

Such norms can become key strategic assets for copytesting companies since they are the source of much of their research and development (R&D) and can form the basis for their claims of predictive validity. Category-specific norms are regarded by client advertisers as being more valuable than the overall normative databases of the company. However, the norms that are most predictive are those that are specific at a brand level. Because a norms-based system can lead to years of loyal use of a given copy-testing technique by a client, norms can also be central to the basic business model of ad testing companies.

Ad Testing Formats

There are a variety of formats of rough advertisements used in testing. Some of the more predominant forms are the following:

- Animatics
- Photoboard/storyboard-based simulations
- Live-action roughs

Animatics use cartoons to simulate the visuals of the advertising. Photoboards use early or stock-based photos to re-create the intended visuals. Live-action roughs use live actors and sets but attempt to compromise on the expenditures that would be used to produce the final commercial or commercials.

Usually, when advertising is quantitatively tested in rough form, the intent is to screen a larger number of commercial alternatives down to a smaller number of commercials. Once this screening test is conducted, the winning commercial or commercials will typically be produced in final production and then retested. All of these tests, at both stages, qualify as ad pretests, inasmuch as all such tests are still conducted prior to going on air.

Objectives of Pretests

The basic purposes of ad pretests are threefold:

- to select the optimal commercial for actual media placement,
- to determine whether specific components of the ads need to be altered or improved, and
- to determine whether the optimal advertisement is likely to perform in a superior manner to advertising that is currently on air. (NOTE: This last objective is obviously not relevant in the case of pretests being conducted for new products, which have not as yet received any live advertising exposure.)

However, because of severe time pressures in today's ad testing environment, ads are often tested after they have already begun to run on air

or are not tested at all. An unknown question is the proportion of advertising campaigns that are receiving airtime/significant media but have not been exposed to consumers for test purposes prior to running on air. If this trend were to continue, the advertising research industry would receive increasing levels of justifiable criticism for a lack of standards of accountability.

Post On-Air Ad Tracking

Once the final advertising execution has been tested and possibly refined and retested, it is placed in the media. Decisions are simultaneously being made by the client with respect to how much money will be spent on the ad campaign, whether the campaign should consist of one execution alone or multiple versions or poolouts, whether the poolouts themselves should also be tested, how the advertising campaign budget should be allocated across the media alternatives, how the budget should be spent across time periods, and whether the versions of the advertising campaign to be placed in alternate media vehicles should also be tested.

For example, it might be the intent of a given campaign to be aired with a certain proportion of the budget allocated toward a TV campaign, say, 60%, but that the other 40% be spent on a mix of print ads, radio ads, and/or online banner ads. Each of these media alternatives requires slightly different ad testing techniques, to be discussed below.

Once these decisions are made and the campaign is aired, another research project may be undertaken, called post on-air ad tracking.

Pretests are individual research projects conducted at a given point in time. However, ad tracking projects attempt to measure the effectiveness of the ad campaign over the course of time. A typical ad tracking project is an annual contract, calling for interviews to be conducted over the course of a year. In the case of an ad tracker for a new product, or if an ad tracking project has not already been set up to monitor the advertising in the client's category, the project may include both "pre" and "post" interviews, conducted both before and after the ad begins to run on air.

Projects may be conducted using a variety of data collection methods, such as telephone, online, mall intercept, or mail interviews. Over the past 20 years or so, there has been a shift from periodic or wave-based interviews for such projects to continuous interviewing. In a wave-based project, a set of interviews might be conducted prior to the onset of advertising, perhaps consisting of 300 to 800 category users, conducted the month prior to the beginning of the ad campaign. Additional interviews would then be conducted at periodic intervals following the campaign's onset, perhaps at 3 months following the campaign's onset and 6 months, 9 months, and 12 months following the campaign's airing. The design of the project would depend on a variety of factors, including the expected media weight, as well as the extent to which it is evenly spread over the months of the campaign, front-loaded toward the early months, or back-loaded toward the end of the year. The seasonality of the category and brand is another factor affecting the research design.

The potential problems with wave-based ad tracking projects are that it is impossible to predict how or when competitors may be simultaneously changing their media spending or strategies and that the actual peaks or valleys in the campaign's effectiveness could easily occur during the periods in which the interviews are not being conducted.

Continuous ad trackers include an interviewing design that calls for a steady stream of interviews being conducted throughout the 52 weeks of the year. In this way, the data can be shown in the form of weekly or monthly data and accumulated in the form of moving totals on the key metrics, such that the peaks and valleys of the campaign's effectiveness can be matched to the actual points in time of spending. The typical sample size in an ad tracking project might be an annual sample of 1,000 to 3,000 interviews, again depending on such issues as anticipated media weight and likely competitive activity.

The Key Quantitative Metrics

Attention

Because it is transient, attention is difficult to measure. Attention can be inferred by asking the consumer to indicate which visual elements

from an ad he or she recognizes after the ad is removed from view. To measure attention by recognition requires a system for selection of the ad content and a means of interpreting the pattern of responses obtained. A simpler means is that of passively observing where a consumer's eyes are directed when he or she is exposed to advertising by tracking with a video camera and a reflected light beam. The measure of attention involves both how much time is devoted to a particular element of an ad and the sequence in which an ad's elements are viewed. When measures of eye movement are combined with specific questions of consumers on the more traditional measures, an effective and insightful analysis can result (Weinblatt, 1999).

Recall of Brands and Sales Points

For a given advertisement to be effective, consumers in the target audience, such as buyers of the given product category, need to not only have their televisions tuned to the station airing a given commercial but also need to see the commercial and to connect the commercial to the brand being advertised. The visuals in a commercial may easily register with consumers and can sometimes even entertain consumers. But if there is no registration in the consumer's mind of the brand being advertised, millions of dollars of the client's advertising budget can be wasted.

Because of this fact, advertisers have been interested in determining the extent to which the consumer is able to accurately recall the brand (i.e., connect the advertising with the brand). Over the years, many methods of measuring brand recall have been developed, and several companies have emerged with varying approaches to this measurement.[1]

There are interactions between the media and the message communicated. For example, a series of ads that all run concurrently on television is called a pod. Experimental data suggest that ads placed toward the end of a pod of commercials may generate higher brand and sales message recall. However, other data suggest that radio commercials may benefit from placements toward the beginning of a pod.

One of the central issues with respect to the measurement of recall involves the approaches used by the research companies to simulate the experience in which consumers are actually exposed to advertising in their homes. Some companies may recruit consumers and then expose them to the test advertising in a central location, such as a mall or in a hotel ballroom. Others may send videotapes to consumers at home. And still others may ask consumers to look at downloaded advertising over the Internet.

Some companies may expose consumers to a "clutter reel" of commercials, including a series of test and nontest commercials. Some companies may also embed commercials in an actual pilot program, while others may not feel that this level of simulation is necessary.

In addition, the questions used to measure brand-name registration vary from company to company and from medium to medium.

Some companies may ask the consumer to name the brand advertised, once she or he says yes to a question about having seen any advertising for a brand in a given product category. Others may provide a list of brands to consumers and ask them to name the advertised brand. Others may show advertising to consumers with the brand name omitted and ask consumers to identify the brand. And still others may prepare a paragraph describing the advertising's visuals and ask for the consumer to name the advertised brand. Put another way, some companies use methods of unaided ad "recall," while others prefer methods more closely related to ad "recognition."

The industry does not have an agreed-upon point of view about which method or methods of measuring brand-name recall are superior. Nevertheless, there is general agreement that forging a connection between the advertising and the consumer's ability to connect the advertising with the brand being advertised is a critical component to effective advertising.

The communication of advertising content, or "sales points," is another common area of measurement. Here, however, there is less industry agreement about the necessity, or even feasibility, of its measurement. For example, while many advertising campaigns have as their objectives the communication of specific and rational claims or content, other campaigns may have as their objectives the communication of more subtle or purely imagery-related communication. The creation of an emotional connection between the

consumer and the brand may be the objective, rather than the communication of specifically delineated claims. Sales might be maximized by a more rational ad campaign for one brand, while a different brand might well be better served by an emotional campaign, containing little if any real product claims or supporting evidence.

Message recall is also measured using a variety of questions, and no industry consensus exists with respect to the proper questions to ask. Still, when two campaigns are equivalent to one another on other metrics and when a rational approach is deemed appropriate, there is general agreement that the campaign that does a better job of communicating relevant sales messages is more likely to achieve its sales objectives.

Persuasion

Persuasion refers to the ad campaign's ability to increase the likelihood of the advertising to motivate target consumers to actually purchase the brand being advertised. It is often regarded as the other most important measure of advertising effectiveness, along with brand-name recall. Although most copytesting research companies measure persuasion in one way or another, this is an area of intensive R&D and has considerable controversy.

An oversimplified summary of the longstanding philosophical debate on the relative merits of recall versus persuasion goes something like this:

- What good is persuasion if no one can remember your commercial or what brand was being advertised?

versus

- What good is recall if no one is interested in buying the brand, even when they remember your ad?

Some research companies believe that persuasion is a far more important measure of ad effectiveness, while others place more weight on recall-based measures.

As a general rule, recall is more likely to be important for new products and smaller brands—in other words, for those brands at the earlier stages of the AIDA model. And recall is more likely to be a relevant measure for emotional campaigns or those brands that are less dependent on the communication of rational sales benefits.

Persuasion is perhaps more likely to be a critical measure than recall for larger brands and for those brands more dependent on clear and easily communicated rational benefits and claims.

The methods by which persuasion is measured also cover a fairly wide gamut of alternatives. One popular method is the pre/post lottery, which originated with the Schwerin technique. In this method, consumers are asked, prior to their exposure to any advertising, to select a brand from a list of alternatives within the product category. They are told that they will then receive a package of that brand in a basket of incentives for their participation. After ad exposure, they are again asked to select a brand from the same list. The pre-to-post shift in preferences toward the advertised brand is the persuasion measure.

Other methods include a 5-point purchase interest question across brands, again administered before and after ad exposure. In this case, the purchase interest question ranges from whether the respondent will "definitely buy" the brand, to "probably will buy," to "might or might not buy," to "probably will not buy," to "definitely will not buy" the brand. In the case of a new product, purchase interest might only be administered following ad exposure, not both pre- and postexposure.

A third alternative is a pre– and post–constant sum question, in which consumers are asked to "divide your next five purchases" in this category across the brands in the category (NOTE: Another variant is the "next 10 purchase" constant sum.) One advantage of the constant sum question is the information it provides about brand switching since an increase in purchasing for the test brand will be matched by a corresponding decrease in the purchase of another brand.

Emotional Response Measures

As noted in the first section, the use of verbal questioning to study emotional responses introduces task-oriented thought processes, and that has been criticized by creatives and researchers as interfering with the feelings being measured or seeking to have things described that may not be accessible or at a conscious level.

Researchers measuring the effects of advertising through physiological measures believe that they are obtaining a more valid measure because the measurement is totally passive—that is, it requires no conscious activity of the respondent and because the measures are made immediately and continuously throughout the course of the ad exposure. Physiological measurements that are currently being applied to advertising include the following:

- GSR or galvanic skin response
- Heart rate and/or blood pressure changes
- Electroencephalogram (EEG) or brain wave patterns
- Functional magnetic resonance imaging (fMRI)
- Facial electromyogram (EMG), measures of the muscles used in smiling and frowning

Nonphysiological measures that are used to measure attention, feelings, and so on concurrently with the ad exposure include rotating a dial or moving a computer mouse along a prescribed line to register the magnitude of the selected feeling state under study.

Researchers seeking to avoid verbal interference have employed pictorial scales such as happy and sad faces. These measures are often made after the ad exposure and can be inquired with regard to the ad, some part of the ad, or the brand itself.

Diagnostics, Content Analysis, and Integration

The basic purposes of quantitative advertising tests, prior to airing, are threefold:

1. to decide which of multiple advertising executions is superior;
2. to determine whether the advertising, thus selected, is better or worse than the advertising already being run; and
3. to make any necessary improvements in the advertising, in order to optimize its effectiveness, prior to its airing.

To address the third objective, research companies employ a variety of methods and measures. Some use standardized batteries of diagnostic questions. They then compare the responses, for a given ad, against their own or even client-specific normative databases of the answers to these diagnostic questions. These comparisons then provide insights into the ad's strengths and weaknesses, which can then be turned into specific recommendations on areas in the ad that could be improved.

Another approach is to analyze the content of the advertising vis-à-vis its component elements, such as the following:

- The number of brand-name mentions
- The length of time between the beginning of the commercial and the first brand-name mention
- The number of visual package shots of the brand
- The inclusion of a brand logo or symbol
- The number of times such a logo or symbol is shown
- The number of scenes included
- The number of characters
- Whether the ad is in the style of a "lecture," in which the message is communicated directly, via voiceover or character, to the viewer or a "drama," in which case the commercial is a mini-vignette, to be watched by the viewer
- The inclusion of a demonstration of product superiority
- The inclusion of music
- Type of music
- Inclusion of a jingle
- Presence of children/babies

Based on research company–sponsored R&D, evidence exists of the connection between the ad's scores on key measures and advertising content. For example, the literature suggests that there is a systematic relationship between the number of brand-name mentions and the ability of consumers to recall the brand's name and even that the mention of the brand in the first several seconds of the commercial will tend to aid recall. In this way, an analysis of the ad content can aid in diagnosing the ad's ability to generate high scores on recall, persuasion, and/or likeability (Baldinger, 1991; Haley & Baldinger, 1991).

A third area of diagnosis is that of second-by-second measurement of the content of an ad. Consumers can, for example, be asked to react to the commercial while simultaneously moving an electronic dial up or down, signifying the consumer's specific positive or negative reactions to the particular scene or components of an ad. When this "interest trace" then turns above or below a steady-state line, insights can be gained on the specific elements of an ad, including specific actors, tone, or scenes.

Another approach is to use a group of photos of various people of a variety of lifestyles and then match these photos to either the brand or specific advertising. Insights can thus be gained about whether a specific piece of advertising copy is properly communicating the desired imagery.

Sales Effects

The most difficult and, simultaneously, the most important challenge in ad testing is the measurement of the sales effects generated by advertising.

There are many reasons for the fact that this phase of measurement is difficult. For example, survey-based copy tests, by their very nature, cannot include a direct measure of actual sales effects. Intent-to-purchase measures were created to simulate purchase behavior, but there is no statistical evidence of a direct correspondence between "definitely will buy" intentions and actual purchase rates. (NOTE: The literature suggests a rule of thumb that roughly 75% to 85% of those who say "definitely will buy" will actually buy the brand, if they become aware of the brand, and if the brand is adequately available where consumers shop. Similarly, somewhere between 10% and 40% of those who say that they will "probably buy" the brand will actually buy it, under ideal conditions. But there are many intervening factors that cause these percentages to shift upwards or downwards, such as category dynamics, purchase cycles, and competitive activity.)

In addition, many factors, beyond advertising alone, intervene between the exposure of the consumer to advertising and the actual sale of the brand. Media from multiple outlets complicate measurement, as do product quality changes, pricing changes, promotional activities such as coupons, and the retailers' own influences on a brand's sales.

Consequently, while most experimental measurements of the sales effects of advertising involve a real-world exposure of advertising, it is still not a simple task to identify a definitive ad effect.

One system for doing this is the use of in-market split cables. The best-known existing split cable system is Behaviorscan, which is a service of Information Resources, Inc. (IRI). In this system, when advertising effectiveness is the test's objective, every attempt is made to eliminate all test variables, except the advertising execution itself. A typical test might consist of a two-market test, run for 6 months or more. Small markets are part of the Behaviorscan system, such as Eau Claire, Wisconsin, or Pittsfield, Massachusetts. A brand will run Campaign A in one of the cable systems in the test market and Campaign B through the second cable system. These markets were originally chosen based on the fact that the towns had two separate and roughly equivalent cable systems, without significant demographic skews in the subscriber base of either cable system.

Advertising sales effects are measured predominantly using a diary panel in a Behaviorscan test. In each market, thousands of respondents have been asked to fill out diaries that track their purchases across multiple packaged goods product categories. Two subpanels are demographically matched to each other, with one panel being exposed to Campaign A, while the other sees Campaign B. The same pattern of actual programs, day-parts, and actual media spending is used for each campaign in the test. (NOTE: A *day-part* refers to the time of day in which the advertising time has been purchased, such as prime time, from 8 to 11 p.m., daytime, late night, the "early fringe" or early evening newscast period, etc.)

Sales are also measured via the scanning data, provided by the retail outlets in each market to Behaviorscan. However, it is not possible to disentangle and separate the scanning data by campaign since households exposed to both campaigns will shop in a given retail outlet. (NOTE: IRI did an extensive analysis of the effectiveness of advertising, using hundreds of Behaviorscan tests as the database, in the early 1990s. This was called the "How Advertising Works" project. Interested readers should contact either IRI or the ARF for the published findings of these analyses.)

The major drawbacks of using this kind of system for measuring the sales effectiveness of a given ad campaign are the time and expense of performing these tests, as well as the fact that there is a loss of confidentiality in this approach. That is, competitors may be able to read and react to a new campaign, prior to its ability to gain its full national effect, by observing the activities in these well-known test market cities. Consequently, the number of tests that are evaluated via such systems today represents a small minority of all ad campaigns run on air.

Another system used to measure ad effects is simply to run the advertising campaigns on air and then use modeling approaches to disentangle their sales effects, using a combination of national scanning data, national electronic diary panel data, and marketing-mix modeling. These approaches are covered in depth in Chapter 24 of this handbook.

Roles and Responsibilities

The Client

Ultimately, the responsibility to measure and test the effectiveness of advertising rests with the company that is paying for it and has the strongest vested interest in its success: the client. In most companies, the people within the advertiser who are most involved with this process include the brand/product manager, the marketing director/VP, the advertising director (if such a function exists), and/or corporate top management.

The decisions that these management people must make include the following:

- Whether to advertise the brand
- How much to spend on the advertising
- How the budget should be allocated by media type, time of year, time of day, and geography (e.g., nationally vs. individual market "spot buys")
- Whether the pattern of spending should be flighted or continuous (i.e., whether the spending should be concentrated within a given period, then repeated in another concentrated period weeks later, or simply spread evenly over the course of the year)
- Which ad agency to retain for each media type
- Whether to shift ad agencies
- Whether to change the advertising currently running
- Which new ad to run
- Whether to test and/or track the advertising
- How to test the advertising
- How to adjust all of the above over time

Most of these decisions are made as part of a marketing plan, prepared by the brand manager, as approved and adjusted by those in upper management.

The Ad Agency

The successful ad agency serves in a strategic partnership role for the client corporation. Its role is to create successful advertising for the client. However, its role in the creation and evaluation of advertising has shifted substantially over the course of the past 20 years. While many ad agencies serve a strong role in both the creation and evaluation of advertising, many of today's ad agencies are much more focused on the creation of advertising than on its evaluation as to its effectiveness.

Partly as a consequence of this trend, many ad agency research functions are housed within a planning department rather than marketing research. The term reinforces the agency's role in the upfront creation and planning of advertising.

As covered above, many ad agencies serve a central role in the earliest stages of a copy exploratory and will often conduct qualitative or other creatively oriented research projects designed to generate the most productive avenues for further study. The evaluative phase is, however, often turned over to the client.

However, many ad agencies also believe that they must serve a broader and more strategic

role for their clients than simply the creation of revenue-producing advertising.

Some have generated their own databases and theoretical models, which help to describe the distinguishing features of more successful brands, in contrast to those that are likely to be less successful. In other words, some ad agencies serve a helpful and strategic role in the measurement of brand equity. This has been one way in which ad agencies have historically distinguished themselves from other agencies. Many of the larger ad agencies have developed proprietary models and approaches, all of which serve to aid their clients in arriving at an improved understanding of their brands. Young and Rubicam's BAV, or Brand Asset Valuator model, is one of many such approaches to the issue of arriving at an improved understanding of brands (Agres & Dubitsky, 1996).

Many ad agencies also include media planning as an active function, which serves the role of intermediary between the client and the media itself. In other words, the agency's media planning department will offer detailed recommendations to the client concerning how the specific media budget for a given brand should be allocated across networks, geographies, time of year, time of day, specific program types, and programs.

The Research Company

Many research companies specialize in various aspects of advertising research. Invariably, they serve as subcontractors to either the client or the ad agency. Often, a given research company will attempt to serve as a long-term strategic partner to the client organization, in a similar manner to that of the ad agency. In some cases, this means that the research company will introduce multiple services covering the testing of advertising at multiple stages of its development, evaluation, and tracking. Some research companies will also take an active role in the building and maintenance of client-specific normative databases of tests and scores, so that each new ad test can be usefully compared to historical scores. Such databases also can assist the research company in searching for generalizable truths connected to advertising's likely effectiveness and how to improve it.

The Media

The media are the link between the client's advertising and the consumer. Some media concentrate on a given media type, such as TV networks versus publishing companies or radio networks.

Since advertising revenue is the primary source of revenue for media companies, they have an intense interest in the various media research projects that generate the ratings or circulation figures, which drive the prices the media can obtain for their programming.

On the other hand, the media rarely take an active role at the earlier stages of the advertising development and testing process.

Technique Variations by Media Type

The central questions in designing and analyzing an effective experimental test of advertising are as follows:

- How to simulate the environment in which consumers are actually exposed to advertising in a given media
- How many exposures to the ad best relate to the real world
- Whether multiple test ads can be evaluated within the same test session
- Whether real-world commercial avoidance, via such technology as remote controls and/or TIVO, needs to be included in the test
- The relative analytic weight to place on measures of recall, persuasion, and diagnostic areas of inquiry

TV Commercial Testing

Research companies involved in the testing of television advertising use a variety of techniques to simulate the real-world exposure to television and its advertising. Many of these methods have been kept the same for decades by the individual research company. Since there is considerable variation in the methods used, and since no single research company has come to dominate ad testing, it is safe to suggest that no single method of simulating exposure has emerged as being demonstrably superior to another.

For example, some research companies believe that in-home exposure is important. A method to accomplish this is to send VCR tapes to members of prerecruited mail panels. After the consumer watches the tape, which includes a pilot TV program and in which test ads have been embedded, a questionnaire that includes recall questions is administered, perhaps 24 hours after the tape is viewed. Then the consumer is asked to continue watching the tape and is asked pre/post persuasion and diagnostics after the second test ad exposure.

Other research companies recruit respondents into a central location, such as in a mall location or a hotel ballroom. These techniques often include a process of embedding the test ads within pilot programs and attempt to simulate real-world exposure to the ads by implying that the pilot is the central stimulus for research, rather than the ads placed within the program. Again, persuasion and diagnostic batteries are often administered after a second exposure to the test ad or ads, within the same tape. Most such test systems include multiple test ads within the same test sessions, always from noncompeting categories and often for multiple clients. In this way, the expense for running the session is spread across multiple clients, which keeps per project costs down and/or increases the profitability of sessions for the research company.

Another variation is the use of a remote control device, to allow the consumer to switch channels during the session. In this variation, the test ads are contained in all programs, which is called "roadblocking" the ads. This can provide useful diagnostic insights concerning the likeability and probability of generating recall for an ad.

Radio Ad Testing

An interesting way to simulate the exposure to radio ads is to recruit respondents into a central location and then invite them to watch a video simulating a driving experience. A radio plays in the background, which contains the test ads. The consumer is then asked about recall of the radio ads, followed by a reexposure and diagnostic questionnaire.

A variation is to allow the changing of the radio stations, providing a measure of the radio ad's likeability.

Print Ad Testing

Many of the research companies with businesses in TV-oriented copytesting and tracking also offer services in print testing. This is also an area that seems to be shifting rapidly in the direction of Internet-based data collection since the exposure of a print ad on a computer screen offers a natural alternative to traditional methods, which are often slower to administer and far more expensive than might be the case using the Internet.

As in other areas of ad testing, there are a variety of methods by which consumers are exposed to print advertising. In some cases, a test ad is inserted or "tipped in" to a normal magazine issue. Consumers are prerecruited in a central location or mall and are asked to read the magazine in as normal a manner as possible. They are then asked to name the ads recalled. Consumers are then reexposed to the test ad and are asked diagnostic questions.

A variation on this method is to recruit respondents by telephone, then mail a magazine to respondents, in which the test ads have been inserted. Consumers are then called by telephone the day following the magazine's receipt. They are then asked for ad recall, to look at the test ad or ads again, and then complete the diagnostic interview, including a pre/post persuasion measure.

Another approach is the Starch Readership Survey. Starch was one of the originators of large-scale studies of ad readership and effectiveness and still conducts large numbers of print ad tests a year. Starch measures three levels of print ad attentiveness and readership:

1. Read most—the extent to which consumers read half or more of the ad's body copy
2. Associated reader—the extent to which consumers noticed the ad and read enough to recall the brand name
3. Noted reader—the extent to which consumers noticed the ad but did not read the copy

Scores in each area are then compared to Starch's normative database, which can provide valuable insights into an ad's performance. The service is also able to provide insights into the

relative strengths and weaknesses of the ad's headline, photos used, color versus black and white, and specific body copy.

Print ad testing is another area in which eye movement measures can prove valuable, inasmuch as insights can be gained about which particular visuals or words are the most likely to grab the consumer's attention, the relative length of time that will be spent on various elements, and the order in which the ad's elements are likely to be viewed.

Newspaper Ads

There are two kinds of newspaper ads: ads that contain ad content and ads that are simply alerting respondents to a weekly special price of a given brand or item in a local retail outlet. The second kind of ad is rarely tested in advance for its effectiveness but may be evaluated as part of a larger promotional campaign taking place between the manufacturer and the retailer.

The content-oriented newspaper advertising follows the same basic elements as a print ad and, when tested, would follow many of the same procedures as are employed in print tests. However, many of these ads that are run on a transitory basis are changed constantly and are run only on a local or regional basis. For many of these reasons, newspaper advertising is not evaluated with the same frequency as is the case with the other forms of media already mentioned.

Billboard Ads

For many of the same reasons as were just covered in newspaper ads, billboard ads are not always tested for their effectiveness prior to being placed.

However, when a national campaign is contemplated, large amounts of media are being expended, or the billboard campaign is being integrated thematically with a simultaneous campaign being run in other media, the wise advertiser will do a pretest of his or her billboard campaign.

In this case, a simulation of an exposure to a test billboard can take the place of a video, in which the view outside a moving car can be simulated and the test billboard ad can be inserted into the videotape. Such tests are normally conducted within a central location and/or mall environment. The video may contain several ads, some of which are standard "control" ads and some of which may be test ads. After viewing the ads, consumers are again asked for ad recall. A reexposure to the advertising and an interview can then take place on diagnostic measures and persuasion.

This is another medium in which eye movement can provide useful insights into ad content exposure.

Online Ad Testing

Online advertising is still changing rapidly and evolving. Many dollars are spent by advertisers on banner ads. In this case, direct measures of effectiveness can sometimes become available via click-through rates.

However, the early literature on click-through rates indicated that many banner ads were generating low levels of response, with average click-through running at one half of 1% of consumers exposed to the Web page or less. Also, many banner ads were so short and fleeting that little in the way of traditional ad content could be realistically included.

Pop-up ads are one way in which online ads can increase their intrusiveness and click-through rates. However, since such ads are often met with annoyance by consumers, technology is rapidly emerging to prevent pop-up ads.

Online advertising is the youngest of the media, and no single ad testing service has yet emerged with a dominant place in the spectrum of testing systems.

Innovation and Future Challenges

The Internet

The exposure of most forms of advertising to consumers, using the Internet as an exposure mechanism, is likely to revolutionize the current methods by which ads are created, evaluated, and tracked for effectiveness over time. Many of the research companies involved in ad testing are either rapidly adapting their systems to the Internet as a replacement for traditional methods or performing advertising-related research (Spielman & Klein, 2001).

The reasons for this are reasonably clear. It is very likely that Internet-based ad research will be both faster and less expensive than performing such experiments using traditional methods. Large amounts of costs can be saved by removing the interviewer from the process of performing ad tests and by simply replacing the interviewer with a self-administered interview in-home. The time and expense necessary to prepare advertising stimuli, send them to facilities around the country, or to mail paper questionnaires are areas in which the electronic transfer of ad stimuli, questionnaires, and data can all save large amounts of time and money. In fact, most of this technological transformation has already taken place.

One of the only remaining obstacles to this transfer of methods is the extent to which online samples are representative of the population being measured. While online penetration has, for example, risen to the range of 60% to 70% of all U.S. households, there are still questions about the demographic differences between Internet-enabled households and those not yet online. The question of whether the demographic weighting of households would solve all of the possible bias problems is as yet not completely resolved.

But another and perhaps even more important question is the extent to which consumers are not only Internet capable but have fast broadband connections to the Internet, using such technologies as cable or DSL lines. Current estimates are that somewhat less than half of all Internet-capable households have as of yet moved into broadband connections.

This is an important issue because of the time that it takes to download video, as well as video advertising, in an online environment. One solution that research companies are experimenting with is to do other things, such as administering preliminary parts of the questionnaire, while the downloads are taking place.

Still, this is an issue largely restricted to television ads. Print ads might easily be testable in an online environment, as might even radio ads or billboards.

Technology and Commercial Avoidance

Several technological factors have made the testing and evaluation of effective advertising far more difficult over the past decade. These include the following:

- Televisions with remote controls
- Cable and satellite TV systems
- A dramatic increase in the number of broadcast and cable networks
- An explosion in the number of TV channels available
- A dramatic increase in the number of radio networks and channels available, as well as magazines and Web sites
- TIVO and other time-shifting technologies

As a result, it is becoming easier and easier for consumers to avoid exposure to commercials. This has put increasing pressure on the ability of advertising testing systems to adequately replicate a real-world exposure environment. Many companies have huge databases of scores on critical measures, all based on methods developed years ago, which may make the companies resistant to making any dramatic changes in their testing systems. Nevertheless, there are growing needs for methods that incorporate the consumer's reduced attention.

Some companies are now experimenting with new methodological approaches, which attempt to reproduce the consumer's tendencies to shift away from a commercial, either in the middle of the commercial or upon repeated exposures. But this will likely be an area of considerable innovation in the years to come.

An Integrated Research Approach

As we have seen, there is a dramatic difference between the way media are spent in support of a brand and the methods by which advertising is developed, tested, and tracked for effectiveness over time. Advertisers have a strong interest in an integrated marketing and advertising approach. They would very much like to see their television campaign work in a synergistic fashion with their spending, using the same campaign theme across TV, radio, print, and online.

Yet the process of development, testing, and tracking may well involve dozens of separate projects, all using separate methodological approaches, and even separate research

companies. Because research companies tend to develop specific specialties in a given medium or research technique, it is extremely difficult to find a research approach that measures the totality of the advertiser's spending in support of the brand.

This is compounded by the fact that the testing process must stop well short of the ad's true objective: sales. The measurement of sales, using observational approaches to measure actual in-store behavior, is a completely separate research domain than the experimental survey-based school, where ad testing and evaluation reside.

It is this unresolved question—how to find a single multiphase and integrated approach to the measurement of advertising—that remains the ultimate future challenge.

Note

1. It is not the purpose of this handbook to endorse the services of any given company. Several companies specialize in the measurement of advertising in television, while others specialize in print, radio, or billboards. However, several have multiple services, covering different media and/or both ad testing and ad tracking. The major companies involved in ad testing and ad tracking include but are not limited to the following: Ameritest, Bruzzone Research, Communicus, Diagnostic Research, Gallup and Robinson, GfK Custom Research, Ipsos-ASI, Mapes and Ross, McCollum Spielman Worldwide, Millward Brown, The PreTesting Company, Research Systems Corporation, and Taylor Nelson Sofres.

References

Adams, A. J., & Blair, M. H. (1992). Persuasive advertising and sales accountability: Past experience and forward validation. *Journal of Advertising Research, 41,* 20.

Advertising Research Foundation. (1961). *Toward better media comparisons.* New York: Author.

Advertising Research Foundation. (2002). *Making better media decisions.* New York: Author.

Agres, S. J., & Dubitsky, T. M. (1996). Changing needs for brands. *Journal of Advertising Research, 36*(1), 21–30.

Baldinger, A. L. (1991, September 11). *The case for multiple measures: Is one ever enough?* Paper presented at the ARF Copy Research Workshop, New York.

Batra, R., & Ray, M. (1986). Situational effects of advertising repetition. *Journal of Consumer Research, 12*(4), 432–445.

Brandt, D., & Biteau, B. (2000, February). Pre-testing and sales validation? *Admap,* pp. 23–26.

Haley, R. I., & Baldinger, A. L. (1991). The ARF copy research validity project. *Journal of Advertising Research, 40,* 114–135.

Hallward, J. (2005, April). Emotions are the equivalent of first impressions. *Ipsos Ideas,* pp. 3–5.

Johnson, B. (2005, March 14). Ad spending tops 2000 by 25%, hits $141 billion. *Advertising Age.*

Lukeman, G. (1995, February 13–14). *Advertising's role in managing brand equity: What we know from 179 case studies.* Paper presented at the ARF Brand Builder's Workshop, New York.

Ogilvy, D. (1994, April 19). *The importance of the ARF David Ogilvy Awards for research excellence.* Speech given by video at ARF Annual Conference, ARF Information Center Archives.

Puto, C. P., & Wells, W. D. (1984). Informational and transformational advertising: The differential effects of time. In T. C. Kinnear (Ed.), *Advances in consumer research* (Vol. 11, pp. 638–643). Provo, UT: Association for Consumer Research.

Ray, M. L. (1974). Consumer initial processing: Definitions, issues, and applications. In G. D. Hughes & M. L. Ray (Eds.), *Buyer/consumer information processing* (pp. 145–156). Chapel Hill: University of North Carolina Press.

Roberts, K. (1998, March). *We need a new kind of research.* Paper presented at the Advertising Research Foundation Annual Convention: Breaking Away: Setting Your Own Agenda, New York. Retrieved from http://www.saatchikevin.com/talkingit/breaking.html

Rossiter, J. R., Percy, L., & Donovan, R. J. (1991). A better advertising grid. *Journal of Advertising Research, 31,* 11–21.

Spielman, H. M., & Klein, A. (2001, March). Going online: How one research firm adapted to the Internet. *Quirk's Marketing Research Review.*

Weinblatt, L. (1999, October). *The death of print has been greatly exaggerated: Print/TV synergy, the efficient way to advertise today.* Paper presented at the Experts Report on Print Research, ARF Week of Workshops, New York.

Wells, W. D. (1989). Lectures and dramas. In P. Cafferata & A. M. Tybout (Eds.), *Cognitive and affective responses to advertising* (pp. 13–20). Lexington, MA: Lexington Books.

Young, C. E. (2001, March). Researcher as teacher: A heuristic model for pre-testing TV commercials. *Quirk's Marketing Research Review.*

Young, C. E. (2004). *A short history of television copytesting.* Albuquerque, NM: Ameritest/CY Research.

Zaltman, G. (2003). *How customers think: Essential insights into the mind of the market.* Boston: Harvard Business School Press.

28

Measuring Customer Equity and Calculating Marketing ROI

Roland T. Rust
University of Maryland

Katherine N. Lemon
Boston College

Valarie A. Zeithaml
University of North Carolina

The Managerial Problem

Marketing managers and top executives frequently want to know which marketing expenditures to increase, to forecast the profitability among possible marketing investments, and to track them later to determine their financial returns. For example, they want to know how to compare the return from an advertising campaign with that of a service quality improvement program or an interior store redesign with an upgrade to a Web site. These are not easy tasks. In evaluating marketing expenditures, the path to profitability is indirect, arising through improved customer perceptions and attitudes that result in changed buying behavior and higher revenues. Cost reductions are far easier to identify ("I just fired Joe and saved the company $75,000 per year") than to project or document the profitability impact of marketing expenditures.

Customer Lifetime Value

A useful way to approach the marketing accountability problem is to consider marketing's effect on customer lifetime value. If we can measure the impact of marketing on individual customers' lifetime values (future profit streams), then marketing's effects are measurable and accountable. By focusing on marketing's effect on individual customers, rather than the impact of aggregate expenditures, a firm can

create a customer-centered approach to brand management (Rust, Zeithaml, & Lemon, 2004) that involves a customer-centered marketing strategy (Rust, Lemon, & Narayandas, 2005; Rust, Zeithaml, & Lemon, 2000) and ability to evaluate marketing return on investment (ROI; Rust, Lemon, & Zeithaml, 2004).

Customer lifetime value (CLV) is the net present value of a customer's profit stream. Because assignment of many fixed costs is arbitrary (Berger & Nasr, 1998), we often use the net present value of a customer's contribution stream rather than profits. The contribution stream results from future purchases, which in turn result from brand choices that the customer makes. Because some customers spend more than others, they are more valuable to a company.

What Is Customer Equity?

When we aggregate the customer lifetime values of a firm's individual customers, the result is the "customer equity" of the firm. Customer equity is therefore the sum of the customer lifetime values of the firm's current and future customers. From this perspective, the purpose of any marketing expenditure is to increase the firm's customer equity, and any marketing expenditure that does not improve customer equity is money wrongly spent.

Customer Equity and the Value of the Firm

Customer equity is a proxy for the value of the firm (Gupta, Lehmann, & Stuart, 2004). Aside from accounting adjustments for expenditures such as plant and equipment and financial liabilities, the customer equity of the firm is equivalent to the value of the firm. Therefore, documenting the effect of marketing expenditures on customer equity provides a measure of financial return on those investments.

Customer Equity as a Practical Approach to Marketing Accountability

Many companies have adopted customer equity as an approach to marketing strategy and marketing accountability, including at least three of the top 10 *Fortune* 500 companies (IBM, General Motors, and ChevronTexaco). Many other companies have employed these methods, including such well-known firms as Sears, Comcast, and Saks Fifth Avenue. To date, the methods have been applied in more than 30 countries worldwide and on every continent of the world except for Antarctica.

Alternative Approaches to Customer Equity

Direct Marketing/CRM Models

One popular approach to modeling customer equity has been the direct marketing/customer relationship management (CRM) approach (e.g., Rust & Verhoef, 2005; Venkatesan & Kumar, 2004). In this approach, the firm builds a customer database to record each customer's purchases along with marketing activities that have been targeted at the specific customers. The advantage of this approach is that actual customer behavior is being analyzed. The disadvantages of this approach are (a) many firms do not have the appropriate databases, (b) the databases rarely include the customer's choices of competing brands, (c) the set of marketing expenditures that can be analyzed is typically limited to direct mailings and other direct contacts, and (d) we cannot learn why the customer chooses to buy from the firm.

Acquisition Versus Retention Models

In this approach, customer equity is viewed as arising from customer acquisition and customer retention expenditures (Blattberg & Deighton, 1996). The typical assumption is that the firm owns a database that contains customer behavior and customer-firm contacts over time (e.g., Thomas, 2001). The advantages and disadvantages of this method are the same as for direct marketing/CRM models, with the added disadvantage that knowing optimal acquisition versus retention expenditures does not provide precise enough information to examine the impact of particular acquisition or retention expenditures (e.g., should the customer acquisition expenditures be for brand advertising or direct selling?).

Customer Retention–Based Models

Most approaches to customer lifetime value and customer equity begin with the firm's existing customer base and then analyze customer retention (e.g., Bolton, 1998; Gupta et al., 2004; Rust, Zahorik, & Keiningham, 1995). The assumption is that once customers leave, they are gone for good. Recent research has demonstrated that this assumption can underestimate customer lifetime value by as much as 47% (Rust, Lemon, & Zeithaml, 2004). The problem is that customer retention-based models fail to model the possibility of a customer switching back to the original brand, a behavior that happens routinely in many purchase categories, particularly in consumer packaged goods.

Brand-Switching Models

This approach explicitly models the brand-switching pattern of individual customers (Rust, Lemon, & Zeithaml, 2004) using "Markov switching matrices," such as have been used for years to capture brand-switching patterns in consumer packaged goods. For example, an estimated Markov switching matrix might tell us that if Kevin bought Brand A on the prior occasion, then the next time he has a .6 chance of buying Brand A, a .3 chance of buying Brand B, and a .1 chance of buying Brand C. Instead, if he bought Brand B on the prior occasion, then he has a .7 chance of repurchasing Brand B, only a .1 chance of buying Brand A, and a .2 chance of buying Brand C. If we know Kevin's switching matrix, as well as the brand Kevin bought most recently, we can then assess Kevin's future purchase probabilities for all of his future purchases, given that the competitive situation does not change.

Of course, the firm would like to change the competitive situation in a positive direction, meaning that the firm needs to understand what drives the switching matrix and, thus, customer equity. In other words, the firm tries to identify the attributes that will make a difference to customers. Obtaining customer-level evaluations of these attributes typically requires customer survey data similar to the type collected in customer satisfaction surveys (e.g., Rust, Lemon, & Zeithaml, 2004). More detail on this approach is given below.

When to Choose Which Model?

Figure 28.1 provides a simple guide showing which customer equity approach to use depending on the situation the firm faces. If the firm does not have a customer-level database with both customer behavior and customer-level marketing contact information, then the brand-switching-based approach is preferred because the approach uses survey data to explain switching behavior. If a customer-level database does exist, then the choice of model depends on whether the firm faces serious competitors. If competition exists, the brand-switching-based approach is again preferred because it is the only method to model the effects of competition. If competition can be safely ignored, then the next question is whether specific expenditures need to be evaluated, beyond just acquisition versus retention. If they do not need to be evaluated, then the acquisition versus retention model is preferred. If more specific guidance is needed, then we must decide whether all marketing involves direct contact with customers. If so, then the direct marketing/CRM approach is preferred. If not, then the customer retention-based model is preferred because that approach can drill down to the drivers of customer retention (e.g., Kordupleski, Rust, & Zahorik, 1993).

Modeling Customer Equity

For the remainder of the chapter, we will focus on the brand-switching-based approach to customer equity because we believe most companies either lack a complete enough customer database or face significant competition.

Drivers of Customer Equity

To model the brand-switching matrix at the individual customer level, we need to understand what drives customer switching and customer retention. All marketing expenditures or drivers of customer equity can be grouped into three main categories—value equity, brand equity, and relationship equity[1] (Rust et al., 2000). Value equity includes drivers involving quality, price, convenience, and other objective perceptions of the offering. Brand equity, on the other hand, focuses on subjective perceptions such as brand image, brand awareness, and brand ethics. Relationship equity involves factors that increase switching costs that are not subsumed by value equity and brand equity, such as frequent buyer programs and ongoing relationship maintenance activities.

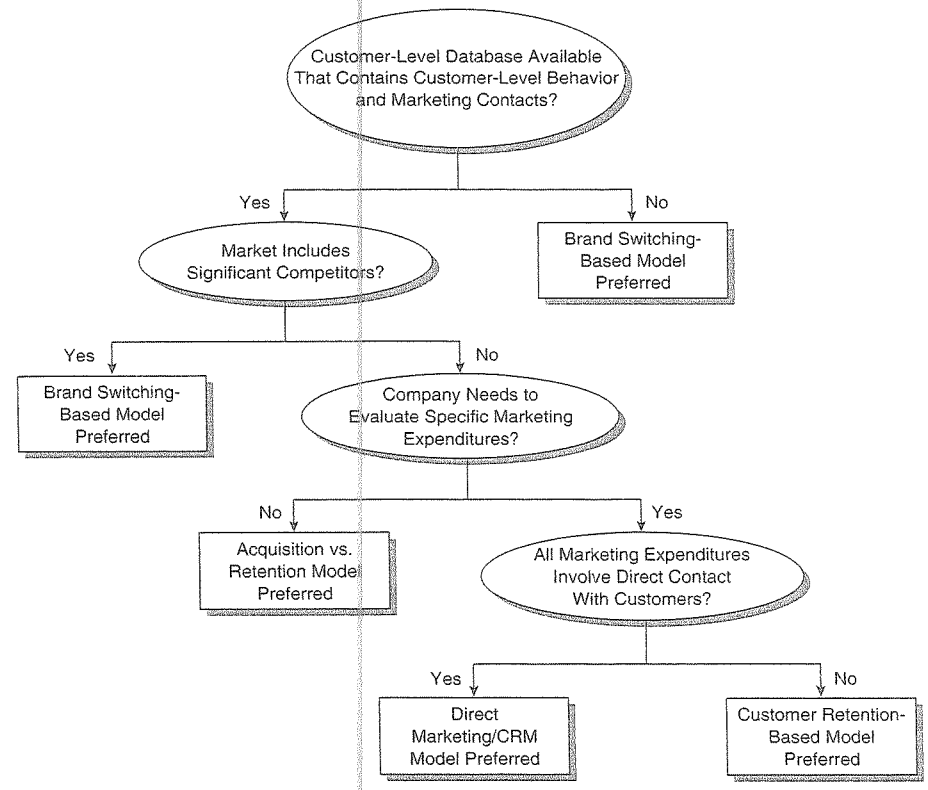

Figure 28.1 When to Choose Which Model

The Chain of Effects

The heart of the brand-switching-based approach to customer equity is a chain of effects model that creates a statistical link from changes in perceptions of the drivers to change in customer equity. The chain is seen at the individual level as follows:

Driver perceptions => Switching matrix => Customer lifetime value

Once this chain is modeled statistically, if the firm can estimate how much it can change the driver perceptions, it can also estimate the impact on customer lifetime value.

The Choice Model

The choice model is conditional on the most recent brand chosen. That is, if Kevin chose Brand A last time, his choice probabilities will be different than if he chose Brand B last time, even if all of his brand perceptions are the same. This condition reflects the effect of inertia on

choice. We model choice based on economic utility, as is standard in the choice literature in economics and as employed in conjoint analysis models in marketing. The customer equity drivers have different weights, reflecting the fact that some drivers are more important than others. If desired, these weights may be estimated at the segment level (using latent segmentation models) or even the individual customer level. A logit regression formulation may be employed to accomplish this (for details, see Rust, Lemon, & Zeithaml, 2004).

The Switching Matrix

The utility of each brand conditional on the previous brand chosen may be obtained according to the following equation:

$$\text{Utility} = \text{Inertia} + \text{Utility from drivers} + \text{Random error.}$$

The inertia term enters the equation only for the choice alternative that was selected most recently. This reproduces the pattern that we see in actual brand choice—that a "stickiness" to the choice of brand exists. Based on the utilities, the probabilities of choice may be obtained using a logit formulation. Again, it is important to emphasize that these probabilities are conditional on previous choice and are different for every customer in the sample. An example of a customer's switching matrix can be seen in Figure 28.2 and is described in Box 28.1.[2]

Customer Lifetime Value

Based on the switching matrix, the probabilities of brand choice for all future purchases by each customer may be projected. This, in turn, may be converted to customer lifetime value, assessed using variables such as the average interpurchase time, average quantity per purchase, and the firm's discount rate and time horizon. The firm may then calculate its customer equity by taking the average customer lifetime value from the sample and multiplying it by the number of customers in the market. An example of this calculation is described in Box 28.1.

Data Collection

Data collection involves both internal company information and customer survey data.

Internal Company Information

The company's discount rate and time horizon are the most important information about the company. The discount rate can be seen as the risk-appropriate rate of return that the firm requires for marketing expenditures, and its effect is to attach less weight to future profits. The higher the discount rate, the less the future counts. The time horizon is also important, as the longer the time horizon, the higher the rate of return. Firms often wish to limit the time horizon to a relatively finite time period (e.g., 3 years) over which they are confident that the market will not change dramatically. A shorter time horizon also limits the amount of time in which competitors might respond, making response within the span of the time horizon less likely, thus making management more confident in the model's projections. For example, in highly uncertain and turbulent markets, the discount rate a firm will employ may be much higher than in markets that are relatively stable with low uncertainty.

Survey Data

Even if a customer database is available, survey data are essential to identify the underlying causes of customer choice. First, we must discover the key attributes that customers use to determine their brand choice behavior. Next, we ask consumers to evaluate each major competitor on these attributes. These data can then be matched with customer purchase data from a database. If a database containing consumer purchase data does not exist, survey questions can also include information on the brand purchased most recently, average purchase frequency, and average volume per purchase.

It is a common error to choose the driver items on the survey based solely on exploratory research with customers (e.g., focus groups). In addition to finding out what customers express as important, it is also necessary to eliminate

Box 28.1 The Brand Switching Matrix and Calculating Customer Lifetime Value: Example

Consider a specific example. Suppose we work for Alpha Corporation. Let us consider Anna, one of our customers. If she bought from us on the last purchase occasion, there is a probability (say, .8) that she will also purchase from us the next time. If she bought from Beta Corporation last time, then there is some probability (say, .3) that she would buy from Alpha the next time and some probability (say, .6) that she would again buy from Beta. If there are three firms in the market—Alpha, Beta, and Gamma—then the switching matrix might look like Figure 28.2. The switching matrix gives the probability that Anna will purchase a particular brand, given a particular purchase the previous time.

If we know which brand Anna purchased last (say, Alpha) and what her switching matrix is, then we can project the probability of purchasing a particular brand for any of Anna's future purchases. For example, if Anna bought from Alpha last time, then the switching matrix says that the probabilities of her next purchase are .8 for Alpha, .1 for Beta, and .1 for Gamma. For the following purchase, her probability of choosing Alpha is [(.8 * .8) + (.1 * .3) + (.1 * .1)] = .68. The first term says that she bought from Alpha and then Alpha again. The second term is a purchase of Beta followed by a purchase of Alpha, and the third term is a purchase of Gamma followed by a purchase of Alpha. We can do a similar calculation for the probabilities of Beta and Gamma on the second purchase. Furthermore, we can also do a similar computation for the third purchase, fourth purchase, and so on. Notice that although Anna's brand-switching matrix stays the same, her probabilities of purchase change. This happens because purchase is "sticky," in that purchasing a brand in one purchase typically increases the probability of purchasing again in the next purchase.

		To			
		Alpha	Beta	Gamma	
From	Alpha	.8	.1	.1	
	Beta	.3	.6	.1	
	Gamma	.1	.2	.7	

Figure 28.2 Brand-Switching Matrix for Sample Customer

Knowing the probabilities of purchase is not enough to figure out customer lifetime value. One must also factor in the average purchase rate per unit time, the average purchase volume per purchase, and the expected contribution margin per purchase. We also need to know the company's investment horizon and its discount rate.

To give a simple example of how customer lifetime calculations work, let us assume that the firm's investment horizon is 2 years, and Anna purchases every other year. From Figure 28.2 we know that the probabilities of purchase for Anna this year (given that she purchased Alpha last time) are .8 for Alpha, .1 for Beta, and .1 for Gamma. For 2 years from now, if she purchases Alpha this year, then the probabilities of purchase would again be .8 for Alpha, .1 for Beta, and .1 for Gamma. If she instead purchases Beta this year, then the probabilities 2 years from now would be (this time

(Continued)

Box 28.1 (Continued)

from the second row of Figure 28.2) .3 for Alpha, .6 for Beta, and .1 for Gamma. If she instead purchases Gamma this year, then the probabilities 2 years from now would be (from row 3 of Figure 28.2) .1 for Alpha, .2 for Beta, and .7 for Gamma. By multiplying the probabilities, we can then ascertain that Anna's probabilities of purchase for 2 years from now are (.8 * .8) + (.1 * .3) + (.1 * .1) = .68 for Alpha, (.8 * .1) + (.1 * .6) + (.1 * .2) = .16 for Beta, and (.8 * .1) + (.1 * .1) + (.1 * .7) = .16 for Gamma.

That tells us Anna's brand choice probabilities, but how much are those potential purchases worth? Let us suppose that the firm uses a discount rate of 10% per year, that Anna purchases one unit each time, and that the contribution to profit from a unit is $50. Let us consider Anna's customer lifetime value to Brand Alpha. It has an 80% chance of getting her first purchase, which means an expected contribution of $50 * .80 = $40. Anna's second purchase will be 2 years from now, again with a $50 contribution, but this time there is only a 68% chance that Anna will purchase Alpha, leading to an expected contribution of $50 * .68 = $34. But remember that that purchase is 2 years from now, meaning that the amount needs to be discounted, by a factor equal to $(1 + \text{discount rate})^{-2} = (1.10)^{-2} = .826$. This means that Alpha's expected contribution 2 years from now is worth $34 * .826 = $28.08. Adding up the contributions from Anna's purchases, Anna's customer lifetime value to Alpha is $40 + $28.08 = $68.08.

We can now think about how to calculate the firm's overall customer equity. Customer equity is defined as the sum of the customer lifetime values across all of the market's customers (if we want to get fancy, we can consider both current and future customers). This can be estimated from a representative sample of the customers in the market. To simplify, suppose there are two customers in the sample—Anna ($68.08 lifetime value) and Bill ($48.97 lifetime value). Our estimate of the average lifetime value of Alpha's customers is then ($68.08 + $48.97)/2 = $58.53. Suppose that Alpha's market has 20,000 customers. Then Alpha's estimated customer equity is the average customer lifetime value multiplied by the number of customers ($58.53 * 20,000) = $1,170,600.

items that cannot be connected to improvement processes or managerial actions. For example, one of us conducted a survey with bank customers 15 years ago that identified "warmth" as a key driver of bank perceptions. Management, however, had no actionable way to connect warmth to process improvements. We advocate a "bridge of actionability" criterion in which the firm identifies key areas of management responsibility and only includes these as the domain of attributes identified by customers. A sample survey (from the airline industry) is provided in the appendix.

Selecting the Sample

A limiting factor in many surveys involves sampling only the firm's existing customers. With this approach, no information about reasons customers choose competition are identified, and we do not know how competitors rate on the attributes. We recommend, therefore, that survey data be collected using a probability sample of all customers in the market, including current customers and noncustomers. The firm's own customers may be oversampled, if desired, as long as they are appropriately weighted when conducting statistical analyses.

Methods of Data Collection

Actual purchase behavior data, if they are available, should be used to enhance the information derived from the survey data. If a customer database exists, surveys should be conducted so that the results can be linked to the behavior in the database. If no customer database exists, then self-reported behavioral data by customers need to be used. Self-reported data may suffer from bias, so it is important in such a circumstance to calibrate the survey responses appropriately, often by matching the aggregate survey responses with the aggregate sales or market share figures. For example, if the survey responses suggest that the three firms in a market have the following market shares—Alpha 50%, Beta 25%, and Gamma 25%—but the actual market shares in the market place are as follows—Alpha 40%, Beta 30%, and Gamma 30%—then the responses from respondents may be reweighted to form a close match to known aggregate behavior.

Data Analysis

The data analysis procedures for the brand-switching-based approach are complex, and the reader is referred to Rust, Lemon, and Zeithaml (2004) for details. This section provides a conceptual overview of the requirements.

Estimating the Choice Model

Estimation of the choice model is similar to that of a multinomial logit model with several important differences. Each customer in the sample has a utility expression for each competing brand in the market, with utility comprising a linear combination of the customer equity drivers (the respondent's attribute ratings), plus an inertia dummy variable term (that takes on a nonzero value when the brand evaluated is the one most recently purchased). The dependent variables are the probabilities of next purchase for each brand, obtained either by tracking the respondents' actual behavior or from self-reports on the survey. The dependent variable is, therefore, not the usual 0-1 in a typical logit application. Even so, many standard logit software packages such as NLOGIT can perform the estimation. Because the customer equity drivers are likely to be highly multicollinear, we regress on the largest principal components, rather than the original variables. The original variables' coefficients and their significance can then be obtained from the logit coefficients and loadings (Rust, Lemon, & Zeithaml, 2004). This enables the importance of each of the drivers to be evaluated.

Generating CLV and Customer Equity

The estimated coefficients enable the firm to project the customer lifetime value of each customer to each firm in the market. Averaging this across customers in the sample and multiplying by the number of customers in the market yields each firm's customer equity. The same coefficients enable the firm to project changes in CLV and customer equity that would result if customer perceptions of any of the drivers changed. Thus, if changes in customer ratings of the drivers can be predicted, the impact on customer equity can be simultaneously predicted. A specific example of the effect of such a change in a driver on customer equity is described in the "Issues in Implementation" section.

Generating Strategic Insights

Customer Equity Share

Marketers are accustomed to looking at market share as the primary measure of strategic health, but we argue that this is shortsighted. Market share is a snapshot of the firm's current standing but does not provide enough information about the future. For example, if all of the firm's customers are 90 years old, the future of the firm 5 years from now will look much less favorable. A better measure of marketing success is the customer equity share, defined as the firm's customer equity divided by the sum of the customer equities of all of the firms in the market. If the firm has a good current market share but a poor customer equity share, then the firm is on the decline. Conversely, if a firm has a poor current market share but a good customer equity share, the firm is clearly on the rise. The customer equity share counts because it represents the net present value of future cash flows, an appropriate measure of the value of the firm.

Comparative Driver Performance

Comparing a firm's performance on the customer equity drivers with those of its competitors is useful. Such analyses have been done for years by companies with customer satisfaction measurement programs (e.g., Fornell, Johnson, Anderson, Cha, & Bryant, 1996). Drivers on which the firm has a competitive advantage can then be leveraged. For example, if

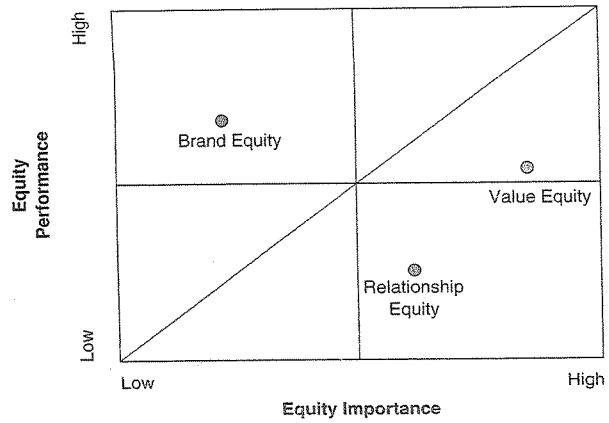

Figure 28.3 Performance-Importance Analysis

a firm dominates the competition on a certain aspect of quality (say, delivery time), this competitive advantage can be the focus of firm communications to current and potential customers. Alternatively, drivers on which the firm trails the competition can be viewed as opportunities for improvement and competitive benchmarking.

Performance-Importance Analysis

Another helpful method that we can borrow from customer satisfaction measurement is the performance-importance analysis. In this approach, a quadrant map is built in which performance (high-low) is mapped against importance (high-low) for each of the drivers. The low performance–high importance quadrant is the most important quadrant for companies to focus on because drivers that are improved in this quadrant make the largest impact on overall assessments. Competing brands can be mapped on the same map, providing a strategic sense of the firm's position. When the firm rates lower than competition on drivers with high importance, the firm needs to focus strategically on these areas. An example of such a performance-importance map appears in Figure 28.3. In this example, we can see that the firm needs to focus on value equity and relationship equity, which are of high importance to the customer and on which the firm is not performing well.

Calculating Marketing ROI

Projecting the ROI of a Proposed Expenditure

Every marketing expenditure should be targeted to improve at least one customer equity driver. The level of improvement expected can be established based on managerial judgment, experience, simulated test markets (Clancy, Shulman, & Wolf, 1994), or full test markets (Simester, Hauser, Wernerfelt, & Rust, 2000). Given the targeted amount of improvement, the impact on CLV and customer equity can then be projected as discussed in the "Data Analysis" section. Given the cost of the marketing expenditure (the discounted net present value of any cost stream), it is then possible to project the return on investment that will result from the expenditure according to the following simple formula:

$$\text{ROI} = (\text{change in customer equity} - \text{marketing expenditure}) / (\text{marketing expenditure}).$$

Projected Return on Relationship for Salesperson Investment

Discounted Expenditure	$180,000
Improvement in Relationship Perceptions	.1
Improvement in Relationship Equity	10%
% Improvement in Customer Equity	2.07%
$ Improvement in Customer Equity	$234,846
Return on Relationship	**30.5%**

Projected Return on Advertising

Discounted Expenditure	$200,000
Improvement in Brand Attitude Perceptions	.1
% Improvement in Brand Equity	16.0%
% Improvement in Customer Equity	5.72%
$ Improvement in Customer Equity	$256,389
Return on Advertising	**28.2%**

Figure 28.4 Comparison of Marketing Investments

This formula enables competing marketing expenditures to be evaluated on the same basis and also allows marketing expenditures to be compared to other corporate investments. This ability to evaluate, compare, and justify marketing expenditures strengthens marketing's position in the boardroom (see example in Figure 28.4).

Evaluating ROI After the Fact

The traditional way to evaluate the ROI of a marketing expenditure is to conduct a before-after experiment. That is, the firm observes the sales before the marketing expenditure, implements the expenditure, and then observes sales after the expenditure. The difference between the before-and-after periods shows the impact of the expenditure, and ROI can then be calculated. This general analytical approach can be extended by the use of control groups (e.g., regions or stores that do not receive the expenditure) or by applying econometric time-series models if there is a stream of expenditures over time (as in advertising, for example).

This analysis yields results that are not completely accurate. The actual ROI stems from the sales that have occurred in the current period, compared not only to a previous period's sales but rather to the sales that would have occurred in the current period had the marketing expenditure not been made, plus similar comparisons (discounted appropriately) in all future periods to which the expenditure applies.

Although not widely realized, the observation of actual changes in sales is not always the most accurate way to measure marketing effectiveness. Suppose, for example, that many extraneous variables have a major impact on sales but that it is impossible to control for many of these variables statistically. Given the messy nature of the real world, this situation is the rule rather than the exception. In these situations, it can be

shown mathematically that comparing the customer equity driver ratings and then projecting the difference in customer equity yields a more accurate estimate of realized ROI than comparing actual sales.

To estimate this, the firm observes the actual driver improvement, inserts that into each customer's utility equation to produce revised CLVs, which then result in an updated customer equity for the firm. The realized marketing ROI can then be calculated as in the "Calculating Marketing ROI" section.

Issues in Implementation

Decision Support Systems

The statistical nature of the model enables the construction of "what-if" simulators that can be used to explore the ROI of marketing expenditures, either before or after the fact. We have constructed such simulators using both spreadsheet models (which get very large and difficult to construct) and dedicated software packages. The advantage of using a simulator is that the company can test out many possible expenditures and then choose only the projects that are likely to produce an attractive ROI. A well-designed customer equity decision support system also provides strategic guidance with respect to the firm's competitive position on each of the customer equity drivers. Figure 28.5 show a sample screen capture from a software package that we developed, using actual but disguised data. In the figure, we see that the airline industry manager is exploring a potential expenditure to improve cabin service quality from an average of 3.6 on a 5-point scale to 3.8. The expenditure costs $100 million up front, plus $10 million per year to maintain. Given a discount rate of 10% and a time horizon of 3 years, the simulator shows that the improvement would result in an ROI of 16%. One can perform sensitivity analysis on such estimates as well. For example, if the same expenditure actually were to instead improve cabin service quality from an average of 3.6 to 3.9 (a 0.3 improvement), then the improvement would generate an ROI of 74.6% rather than the 16% noted above.

Tracking Customer Equity

Markets are dynamic and the competitive environment changes regularly, implying that the competitive situation may change over time. For this reason, customer equity and its drivers need to be tracked over time. We recommend that a company update its analysis at least twice per year to monitor competitive trends, identify emerging threats and opportunities, and gauge the progress of marketing initiatives. Such tracking studies are common in marketing, especially in areas such as customer satisfaction and brand equity, both of which can be subsumed by the customer equity approach.

Conclusions

Firms are increasingly realizing that their financial health is based on the value of their customer base. Customer equity, the sum of the lifetime values of the firm's current and future customers, is the best measure of the value of the firm's customers and is a good proxy for the total value of the firm. An increasing number of leading firms are adopting methods for measuring and analyzing customer equity, using them to make their marketing efforts financially accountable.

Different approaches to customer equity are appropriate to different circumstances. Direct marketing/CRM models and acquisition versus retention models typically require the firm to have a customer database that includes all purchase information and all marketing contacts with each customer. For those firms with these databases and for whom competition is not an important issue, the models that emerged from the direct marketing paradigm are preferred.

For firms without such databases or for which competition is important, a brand-switching-based approach to customer equity is preferred. The brand-switching approach employs customer surveys to identify the underlying reasons that customers choose to stay with their existing brand or switch to another. These reasons or drivers can be classified into three main groups—value equity drivers that involve more objective evaluations, brand equity drivers that involve subjective or emotional evaluations, and relationship equity

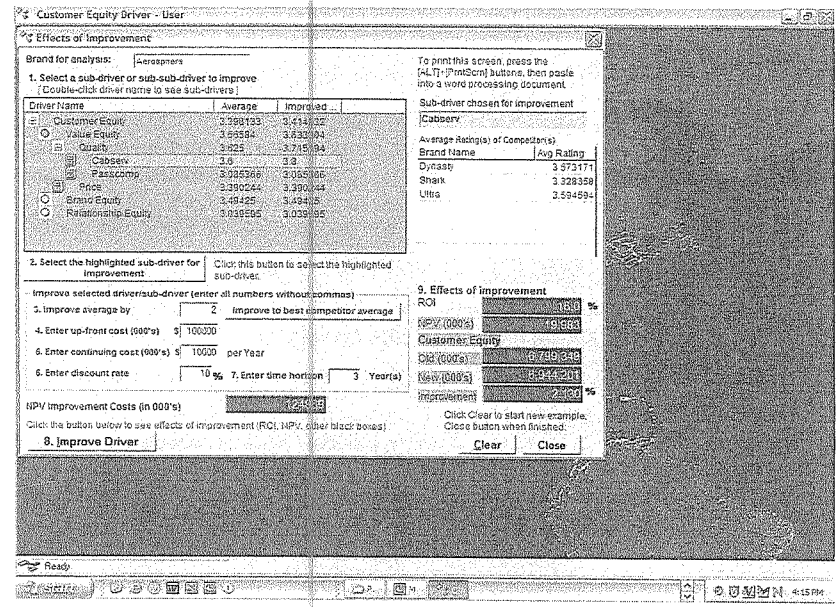

Figure 28.5 A Decision Support System for Return on Marketing

drivers that involve relationship-building factors that increase switching costs.

Building a statistical model of customer equity enables the firm to project the potential impact of improving customer perceptions of the drivers of customer equity. By combining this impact with the cost of effecting the improvement, the firm can project the ROI that results from a marketing expenditure. The same approach may be used after the fact to determine the ROI that has resulted from a marketing expenditure.

In terms of implementation, survey-based customer equity models may be thought of as subsuming all of marketing into a customer satisfaction-like tracking system. Survey data may be linked to a customer database, if available, to provide further calibration and validation of stated or predicted customer behavior.

Firms have long wanted to make their marketing programs financially accountable, and the customer equity approach makes this perpetual wish practical to even medium-sized companies. The result is that marketing can justify its expenditures more effectively, thereby assuming a more prominent position in the corporate boardroom.

Appendix

Example Survey Items (Airline Survey)

Here are some examples of survey items that might be used to measure customer equity and its drivers. These items are from the survey that we used to analyze the airline market:

Market Share and Transition Probabilities[3]

1. Which of the following airlines did you most recently fly (please check one)? (*provide a list of relevant airlines, including the focal firm and key competitors*)

2. The next time you fly a commercial airline, what is the probability that you will fly each

of these airlines? Probability (*please provide a percentage for each airline, and have the percentages add up to 100%*)

Size and Frequency of Purchase

3. When you fly, how much on average does the airline ticket cost?

   _____ less than $300
   _____ between $300 and $599
   _____ between $600 and $899
   _____ between $900 and $1199
   _____ between $1200 and $1499
   _____ between $1500 and $1799
   _____ between $1800 and $2099
   _____ $2100 or more

4. On average, how often do you fly on a commercial airline?

   _____ once a week or more
   _____ once every two weeks
   _____ once a month
   _____ 3–4 times per year
   _____ once a year
   _____ once every two years, or less

Value Equity Drivers

5. How would you rate the overall quality of the following airlines? (5 = Very High Quality, 1 = Very Low Quality)[4]

6. How would you rate the competitiveness of the prices of each of these airlines? (5 = Very Competitive, 1 = Not at All Competitive)

7. The airline flies when and where I need to go. (5 = Strongly Agree, 1 = Strongly Disagree)

Brand Equity Drivers
(5 = Strongly Agree, 1 = Strongly Disagree)

8. I often notice and pay attention to the airline's media advertising.

9. I often notice and pay attention to information the airline sends to me.

10. The airline is well known as a good corporate citizen.

11. The airline is an active sponsor of community events.

12. The airline has high ethical standards with respect to its customers and employees.

13. The image of this airline fits my personality well.

Relationship Equity Drivers
(5 = Strongly Agree, 1 = Strongly Disagree)

14. I have a big investment in the airline's loyalty (frequent flyer) program.

15. The preferential treatment I get from this airline's loyalty program is important to me.

16. I know this airline's procedures well.

17. The airline knows a lot of information about me.

18. This airline recognizes me as being special.

19. I feel a sense of community with other passengers of this airline.

20. I have a high level of trust in this airline.

NOTES

1. Relationship equity was originally referred to as "retention equity."

2. This example is adapted from Rust et al. (2005, pp. 142–145).

3. The headings in this appendix are for explanatory purposes and would not be read to the respondent.

4. The list of relevant airlines (the focal firm and key competitors) is provided to respondents for Questions 5 to 20.

REFERENCES

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing, 12,* 17–30.

Blattberg, R. C., & Deighton, J. (1996). Manage marketing by the customer equity test. *Harvard Business Review, 74,* 136–144.

Bolton, R. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science, 17*(1), 45–65.

Clancy, K. J., Shulman, R. S., & Wolf, M. (1994). *Simulated test marketing.* New York: Lexington Books.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing, 60,* 7–18.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41,* 7–18.

Kordupleski, R., Rust, R. T., & Zahorik, A. J. (1993). Why improving quality doesn't improve quality. *California Management Review, 35,* 82–95.

Rust, R. T., Lemon, K. N., & Narayandas, D. (2005). *Customer equity management.* Upper Saddle River, NJ: Prentice Hall.

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68,* 109–127.

Rust, R. T., & Verhoef, P. (2005). Optimizing the marketing interventions mix in intermediate-term CRM. *Marketing Science, 24*(3), 477–489.

Rust, R. T., Zahorik, A. J., & Keiningham, T. L. (1995). Return on quality: Making service quality financially accountable. *Journal of Marketing, 59,* 58–70.

Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2000). *Driving customer equity: How customer lifetime value is reshaping corporate strategy.* New York: Free Press.

Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2004). Customer-centered brand management. *Harvard Business Review, 82*(9), 110–118.

Simester, D. I., Hauser, J. R., Wernerfelt, B., & Rust, R. T. (2000). Implementing quality improvement programs designed to enhance customer satisfaction: Quasi-experiments in the United States and Spain. *Journal of Marketing Research, 37,* 102–112.

Thomas, J. S. (2001). A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research, 38,* 262–268.

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and optimal resource allocation strategy. *Journal of Marketing, 68*(4), 106–125.