

Evaluation of Effectiveness



It's a Winner

Title:

Client:

Agency:

the “canned spam” laws, everyone on the list has opted in and asked for e-mail promotions and information. This is a perfectly qualified database of customers—they own our product and raised their hands for more information.

As we headed into the selling season for 2009–2010 season passes, the marketing department was charged with ensuring that (1) our loyal pass holders stayed loyal and renewed—locking them in for the season, (2) previous pass holders returned, and (3) individuals with competitor passes converted to the Colorado Pass.

Once we had secured a healthy base of season-pass holders we had to focus on generating revenue for the various lines of business within the company (lodging, ski school, rentals, dining). The challenge for us was that it had been a particularly difficult landscape. The recession was in full swing and we had to contend with cautious consumers, elevated gas prices (something the drive market was sensitive to), and a relatively low snow season (this group lives for snow days).

Beyond encouraging our pass holders to renew their passes, what could we do to take advantage of that pass holder database?

The concept of the Colorado Pass Club (CPC) was introduced as the 2006–2007 ski season wound down. The objectives for the CPC were to distance the Colorado Pass from competitive passes by offering elevated benefits. In particular, we wanted to create a sense of community among Colorado Pass holders. Moving forward, we wanted to continue to drive awareness of the inclusive and exclusive aspects of the pass program.

To further the sense of community and ownership within our pass holder group we developed the tagline “It’s Why I Live Here.” This tagline strategically positions our resorts and having a pass with other iconic Colorado experiences. Skiing, and being outdoors, is an integral part of living in Colorado. We felt that this statement spoke to a broad market, tapped into sense of pride and exclusivity about living in Colorado, and complemented the CPC efforts well.

Our CPC program included biweekly e-mails full of promotions, offers, and discounts exclusively for pass holders. Promotions ranged from dining and event tickets to deeply discounted lodging, ski tune-ups, and discounts on gear. We also gave away tickets to sporting events, free mountain lunches, gas cards, and tickets to music events. Called “Random Acts of Pass Holder Kindness,” these offers were diverse enough to create a strong response and a lot of interest and to elevate the perceived value of the season pass.

Many of our efforts were also supported by radio spots and a special CPC website that recapped the contents of our biweekly e-mails. Slowly these e-mails started creating buzz in the pass holder community, and soon they were being forwarded to friends and family.

Not only did the CPC e-mails generate interest and excitement with Colorado Pass holders, they generated trackable revenue—everything from lodging purchases to ski school lessons, to equipment rentals and, of course, season pass renewals. In fact, our e-mail programs were one of the strongest performing channels for our promotions!

We decided to leverage the passion around our season passes and launched a referral program, rewarding skiers for referring their friends to buy a pass—furthering our intent of building a community. We were able to test various versions of e-mails—we learned that simpler was better and started to reframe offers by coming up with a few high-impact offers rather than several small gestures.

As the program continues to evolve, we have learned quite a bit about our pass holders and have begun to not only offer them great benefits and expressions

CHAPTER KEY POINTS

1. Why and how is marketing communication evaluation conducted?
2. Can you list and explain the stages of message evaluation?
3. What are the key areas of media evaluation?
4. How are IMC tools, campaigns, and programs evaluated?

The Colorado Pass—The Coolest Club on the Slopes!*

Here's the problem: How do you not only maintain share of market but increase business for a ski resort in a competitive landscape? I work for Vail Resorts and that's one of our biggest challenges.

Although the business from the destination market is important to the financial health of the company, so are season passes. Season-pass sales are a significant contributor to the bottom line and provide revenue stability by locking in revenue early in the ski season. Lift ticket revenue is 35 percent of all company revenue; season passes represent approximately 33 percent of that lift ticket revenue. In fact, most season passes are purchased before the season even begins. However, season-pass purchases are conducted in a very competitive environment because our resorts go up against other Colorado resorts to secure the largest share of the 5 million skiers in the local market.

Our company is in a unique position to attract skiers and snowboarders because of its five mountains—Vail, Beaver Creek, Breckenridge, Keystone, Heavenly-Lake Tahoe—and Arapahoe Basin (a partner resort). Our largest competitors in Colorado, where competition is the fiercest, include Copper Mountain, Winter Park, Aspen, and Steamboat Springs. The company offers a few different types of season passes, but we'll focus on the Colorado Pass™, our marquee pass.

One of our strengths is our robust customer databases, which contain transactional information (a history of products purchased) and traditional marketing information. The combination of the two permits very strong database segmentation for e-mail and direct-mail efforts. We also track revenue generated from those efforts so that we can assess their ROIs and strengthen the next campaign. We can also tell if a guest has booked a trip or purchased a product through our e-mails or direct mail and can ensure that they aren't served up the same message again—keeping our communication effective and meaningful.

Vail Resorts has more than 500,000 total e-mail subscribers for various programs; more than 70,000 of these are Colorado Pass holders. In accordance with

*This chapter opening story was contributed by Jennifer Wolfe-Kimbell, Senior Marketing Manager at Vail Resorts, Breckenridge, Colorado.

of appreciation, we engage them in a dialogue and use this group for surveys about product attributes and program improvements.

We also designed unique ways to engage them with content by running contests asking why they haven't skied more. When our pass holders responded with stories about job loss, bad equipment, the need for improved ski technique, we responded by finding one pass holder a job, giving gift certificates to replace equipment, and signing them up for free ski school! The pass holder community loved it.

The result cemented the Colorado Pass Club as the most important club for Colorado skiers. You can find out more about what we achieved by turning to the end of the chapter and reading it's a Wrap.

Throughout this book you've read about the effectiveness of advertising. We've emphasized that effective campaigns do more than win awards for creativity. They work hard to achieve the campaign's communication and marketing goals, as does the Colorado Pass Club campaign for Vail. As you will see in this chapter, there are many ways to evaluate the effectiveness of various aspects of an advertising and marketing communication program. Specifically, four categories of work get evaluated in advertising and marketing communication programs: the message execution, the media, the campaign, and the other marketing communication areas and their synergistic effect as part of an IMC program. This chapter discusses the basic concept of the evaluation of effectiveness and then examines those four areas.

IMPACT: DOES IT WORK?

What makes a marcom message effective? The fact that people like it—or the fact that it moves people to some kind of action? The classic campaign for Avis is an example of one that works on both levels—people like it and it drove the company to record sales levels. When advertising works—when marketing communication is effective—it has impact and generates a response of some kind from its target audience. The critical phrase—"as intended"—means there are multiple objectives and, therefore, multiple ways to evaluate the effectiveness of marketing communication campaigns and programs.

Let's just consider advertising to begin with. Many executives feel advertising works only if it produces sales. Syracuse University Professor John Philip Jones, who has written many books and articles on the topic, estimates that of the \$500 billion spent annually on advertising globally, only 41 percent—less than half—produces sales.¹ Jones contends that "advertising must generate an immediate jolt to sales before it can be expected to produce any further effect."² Simon Broadbent, another leading figure in effectiveness research, disagreed with Jones's short-term sales focus and suggested instead that advertising should emphasize long-term brand building.³

Determining advertising effectiveness based on sales can be difficult because of the impact of other marketing mix factors. For example, an article in *Business 2.0* reported that an ad for the Six Flags amusement park was a smash success in viewer surveys. However, because attendance at the company's 31 theme parks

Avis is only No.2 in rent a cars. So why go with us?



We try harder.
(When you're not the biggest,
you have to.)

We just can't afford dirty ashtrays. Or half-empty gas tanks. Or worn wipers. Or unwashed cars. Or low tires. Or anything less than

seat-adjusters that adjust. Heaters that heat. Defrosters that defrost.

Obviously, the thing we try hardest for is just to be nice. To start you out right with a new car, like a lively, super-torque Ford, and a pleasant smile. To know, say, where you get a good pastrami sandwich in Duluth.

Why?

Because we can't afford to take you for granted.
Go with us next time.

The line at our counter is shorter.

CLASSIC

According to *Advertising Age*, the Avis "We Try Harder" campaign was one of the top 10 in advertising history. It was the ultimate simple idea that conveyed a great idea that was truthful and believable. The strategy—to state strongly a self-effacing underdog position—also tapped into people's sympathy for underdogs. The copy defined what it means to try harder (and what a leader in the car rental business should be providing)—clean ashtrays, full gas tanks, wipers that work, washed cars, heaters and defrosters that work—and the lines at the counter are shorter. It also subtly suggested the problems you might find with #1—Hertz. Created by the legendary Doyle Dane Bernbach agency in 1962 when Avis's market share was only 11 percent, the campaign drove a turnaround for the company. Four years later, Avis's market share was 30 percent and the campaign got credit for a 200 percent increase in business.

of appreciation, we engage them in a dialogue and use this group for surveys about product attributes and program improvements.

We also designed unique ways to engage them with content by running contests asking why they haven't skied more. When our pass holders responded with stories about job loss, bad equipment, the need for improved ski technique, we responded by finding one pass holder a job, giving gift certificates to replace equipment, and signing them up for free ski school! The pass holder community loved it.

The result cemented the Colorado Pass Club as the most important club for Colorado skiers. You can find out more about what we achieved by turning to the end of the chapter and reading *It's a Wrap*.

Throughout this book you've read about the effectiveness of advertising. We've emphasized that effective campaigns do more than win awards for creativity. They work hard to achieve the campaign's communication and marketing goals, as does the Colorado Pass Club campaign for Vail. As you will see in this chapter, there are many ways to evaluate the effectiveness of various aspects of an advertising and marketing communication program. Specifically, four categories of work get evaluated in advertising and marketing communication programs: the message execution, the media, the campaign, and the other marketing communication areas and their synergistic effect as part of an IMC program. This chapter discusses the basic concept of the evaluation of effectiveness and then examines those four areas.

IMPACT: DOES IT WORK?

What makes a marcom message effective? The fact that people like it—or the fact that it moves people to some kind of action? The classic campaign for Avis is an example of one that works on both levels—people like it and it drove the company to record sales levels. When advertising works—when marketing communication is effective—it has impact and generates a response of some kind from its target audience. The critical phrase—“as intended”—means there are multiple objectives and, therefore, multiple ways to evaluate the effectiveness of marketing communication campaigns and programs.

Let's just consider advertising to begin with. Many executives feel advertising works only if it produces sales. Syracuse University Professor John Philip Jones, who has written many books and articles on the topic, estimates that of the \$500 billion spent annually on advertising globally, only 41 percent—less than half—produces sales.¹ Jones contends that “advertising must generate an immediate jolt to sales before it can be expected to produce any further effect.”² Simon Broadbent, another leading figure in effectiveness research, disagreed with Jones's short-term sales focus and suggested instead that advertising should emphasize long-term brand building.³

Determining advertising effectiveness based on sales can be difficult because of the impact of other marketing mix factors. For example, an article in *Business 2.0* reported that an ad for the Six Flags amusement park was a smash success in viewer surveys. However, because attendance at the company's 31 theme parks

Avis is only No.2 in rent a cars. So why go with us?



We try harder.
(When you're not the biggest,
you have to.)

We just can't afford dirty ashtrays. Or half-empty gas tanks. Or worn wipers. Or unwashed cars. Or low tires. Or anything less than seat-adjusters that adjust. Heaters that heat. Defrosters that defrost.

Obviously, the thing we try hardest for is just to be nice. To start you out right with a new car, like a lively, super-torque Ford, and a pleasant smile. To know, say, where you get a good pastrami sandwich in Duluth.

Why?

Because we can't afford to take you for granted.

Go with us next time.

The line at our counter is shorter.

CLASSIC

According to *Advertising Age*, the Avis “We Try Harder” campaign was one of the top 10 in advertising history. It was the ultimate simple idea that conveyed a great idea that was truthful and believable. The strategy—to state strongly a self-effacing underdog position—also tapped into people's sympathy for underdogs. The copy defined what it means to try harder (and what a leader in the car rental business should be providing)—clean ashtrays, full gas tanks, wipers that work, washed cars, heaters and defrosters that work—and the lines at the counter are shorter. It also subtly suggested the problems you might find with #1—Hertz. Created by the legendary Doyle Dane Bernbach agency in 1962 when Avis's market share was only 11 percent, the campaign drove a turnaround for the company. Four years later, Avis's market share was 20 percent and the campaign got credit for a 300 percent increase in business.

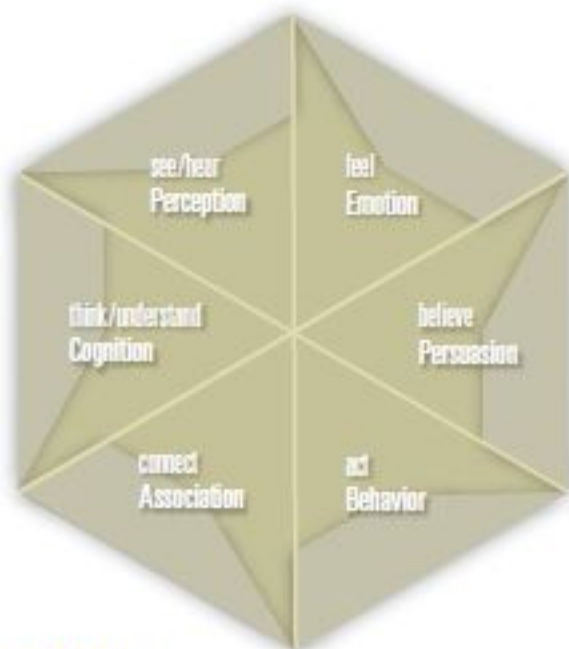


FIGURE 19.1
The Facets of Effects Model

Principle

If you can't measure it,
you can't manage it.

however, is to recognize early in the planning stage the need for a formal evaluation mechanism, as Professor Mark Stuhlfaut explains. Evaluation should be “planned in” to any campaign. Evaluation determines the success of a marketing communication, but it also feeds back into the ongoing brand communication plan. Figure 19.2 illustrates that process, and Professor Stuhlfaut’s *A Matter of Principle* features builds on the idea of a cycle beginning and ending with research.

In addition to intuition and judgment, measurement that tracks consumer response is also needed and can be built into a campaign plan as structured feedback, such as response cards and calls. Often, however, the evaluation effort involves a more formal research project, which also needs to be part of the campaign planning.

Personal insights into what makes an ad great are important, but why do we also need formal evaluation? The first reason is that the stakes in making an advertising misstep are high. By the time an average 30-second commercial is ready for national television, it has cost about \$200,000 in production costs. If it is run nationally, its sponsor can invest several million dollars in airtime. The second reason is advertising optimization, or reducing risk by testing, analyzing, tracking brand performance, and making changes where possible to increase the effectiveness of the advertising. The third reason is to learn what works and what doesn’t—in other words, to identify best practices so a brand’s advertising continues to improve.

Types and Stages of Evaluation

Evaluation is done through testing, monitoring, and measurement. Testing is used to predict results, monitoring tracks performance, and measurement evaluates results. In other words, for major campaigns a sample of the ads is typically tested before they run as a way to predict effectiveness. Ideally, the results of preliminary evaluative research should be available before large sums of money are invested in finished work or in media buys.

As a campaign unfolds, the performance is tracked to see whether any elements need to be changed. Sales may fall, or they may not increase as rapidly as expected. Is the advertising at

fell after the campaign instead of increasing, it must be deemed ineffective. But sales are not the only reason brands advertise: one of the major objectives of advertising is to create higher levels of brand awareness. Perhaps the Six Flags awareness level increased, but the reason people didn’t come to the parks has to do more with the lack of new attractions or maybe an external factor such as an increase in the price of gas. Does that mean the advertising is ineffective?

It’s our view that marketers intend for their messages to accomplish a variety of goals, such as build a brand relationship, recruit volunteers or donations, or entice people to visit a website, which is why the other 60 percent of ads studied by Jones may be effective in their own way. That’s why this book uses the Facets Model of Effects to broaden the way we evaluate message effectiveness.

Evaluating Effectiveness

The most important principle in this chapter—possibly in this book—is a well-known maxim: “If you can’t measure it, you can’t manage it.”

Some evaluation is informal and based on the judgment of an experienced manager. There is always a need for the intuitive analysis of experienced professionals. The important thing,

fault? The results, the actual effects, are measured after the ad or campaign runs. Diagnostic research also is used in all three stages to deconstruct an ad to spot message problems. Four types of research are used in evaluation:

1. **Developmental research** through pretesting estimates the likelihood that an ad idea will work or that one idea is better than another.
2. **Concurrent research** using tracking studies and test marketing monitors the way the campaign is unfolding and how the messages and media are working.
3. **Post-testing research** evaluates impact after the campaign is over or after the ad ran. Postcampaign research encompasses benchmark or baseline studies to gauge movement. These can be research company norms, or they can be based on previous campaigns by this brand.
4. **Diagnostic research** deconstructs an ad to see what elements are working or not working. Researchers who evaluate commercials use frame-by-frame or moment-by-moment analysis to identify strengths and weaknesses in an ad.

Facets: Measuring Responses

Most advertisers would be happy if evaluation could simply tell them how much the advertising contributed to their sales effort. That's difficult for a number of reasons. Factors other than advertising affect sales (e.g., pricing, distribution, product performance, competition), and that makes it hard to isolate advertising to determine its impact. Furthermore, advertising effects tend to be delayed, so it's difficult to link the advertising seen at one time to a purchase made days or weeks later. Exceptions are direct-response advertising and ads carrying promotional offers good for only a limited time.

Usually advertising is measured in terms of its communication effects—the mental responses to a message, which become *surrogate measures* for sales impact. Such factors as awareness of the advertising, purchase intention, preference, and liking all suggest that the advertising message can make a positive contribution to an eventual purchase decision. According to research professionals at Ipsos ASI, one of the largest U.S. providers of advertising pretesting, “Ads work best when they engage viewers’ interest, when consumers enjoy watching them, when they are relevant, and when they tell their story in a unique and interesting way.”⁴ Those are some of the dimensions of effectiveness, but others also are important, as we know from our discussion of how advertising works in Chapter 4.

On the other hand, some ads don't work, and it is just as important to understand why that happens. Some confuse the audience or fail to get attention or connect with consumers. In many cases, people can't remember the brand. In some cases, the ad can even boomerang, an effect mentioned in a feature on antidrug ads in Chapter 4. The same problem was noted with anti-smoking ads when the American Legacy Foundation announced that a study of smoking prevention ads by the tobacco industry actually increased the likelihood that teens would try smoking.⁵

Good evaluation objectives are based on a model of human response to an advertisement that identifies key effects. We developed the Facets Model of Effects in Chapter 4 as a guide for both setting objectives and evaluating effectiveness. Table 19.1 groups the key factors of effectiveness and then matches them to the types of research questions advertisers can use to determine effectiveness.

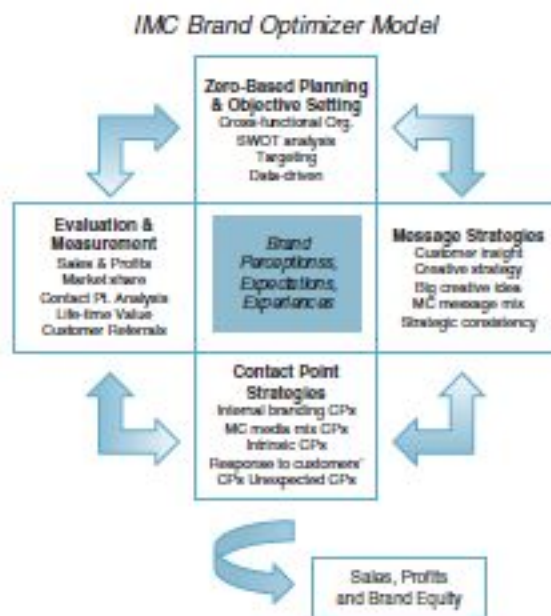


FIGURE 19.2

IMC Brand Optimizer Model

This IMC model created by Professor Tom Duncan illustrates that evaluation is a circular process—IMC plans start by gathering information and move through the various steps in the planning process to come back to the last step in the process, which is again gathering information. This time, however, information is gathered to determine what works and what doesn't. That information feeds back into the process and the organization learns from the results.

Principle

Good evaluation plans, as well as effective professional work, are guided by a model of how people respond to advertising.

fault? The results, the actual effects, are measured after the ad or campaign runs. Diagnostic research also is used in all three stages to deconstruct an ad to spot message problems. Four types of research are used in evaluation:

1. **Developmental research** through pretesting estimates the likelihood that an ad idea will work or that one idea is better than another.
2. **Concurrent research** using tracking studies and test marketing monitors the way the campaign is unfolding and how the messages and media are working.
3. **Post-testing research** evaluates impact after the campaign is over or after the ad ran. Postcampaign research encompasses benchmark or baseline studies to gauge movement. These can be research company norms, or they can be based on previous campaigns by this brand.
4. **Diagnostic research** deconstructs an ad to see what elements are working or not working. Researchers who evaluate commercials use frame-by-frame or moment-by-moment analysis to identify strengths and weaknesses in an ad.

Facets: Measuring Responses

Most advertisers would be happy if evaluation could simply tell them how much the advertising contributed to their sales effort. That's difficult for a number of reasons. Factors other than advertising affect sales (e.g., pricing, distribution, product performance, competition), and that makes it hard to isolate advertising to determine its impact. Furthermore, advertising effects tend to be delayed, so it's difficult to link the advertising seen at one time to a purchase made days or weeks later. Exceptions are direct-response advertising and ads carrying promotional offers good for only a limited time.

Usually advertising is measured in terms of its communication effects—the mental responses to a message, which become *surrogate measures* for sales impact. Such factors as awareness of the advertising, purchase intention, preference, and liking all suggest that the advertising message can make a positive contribution to an eventual purchase decision. According to research professionals at Ipsos ASI, one of the largest U.S. providers of advertising pretesting, “Ads work best when they engage viewers’ interest, when consumers enjoy watching them, when they are relevant, and when they tell their story in a unique and interesting way.”⁴ Those are some of the dimensions of effectiveness, but others also are important, as we know from our discussion of how advertising works in Chapter 4.

On the other hand, some ads don't work, and it is just as important to understand why that happens. Some confuse the audience or fail to get attention or connect with consumers. In many cases, people can't remember the brand. In some cases, the ad can even boomerang, an effect mentioned in a feature on antidrug ads in Chapter 4. The same problem was noted with anti-smoking ads when the American Legacy Foundation announced that a study of smoking prevention ads by the tobacco industry actually increased the likelihood that teens would try smoking.⁵

Good evaluation objectives are based on a model of human response to an advertisement that identifies key effects. We developed the Facets Model of Effects in Chapter 4 as a guide for both setting objectives and evaluating effectiveness. Table 19.1 groups the key factors of effectiveness and then matches them to the types of research questions advertisers can use to determine effectiveness.

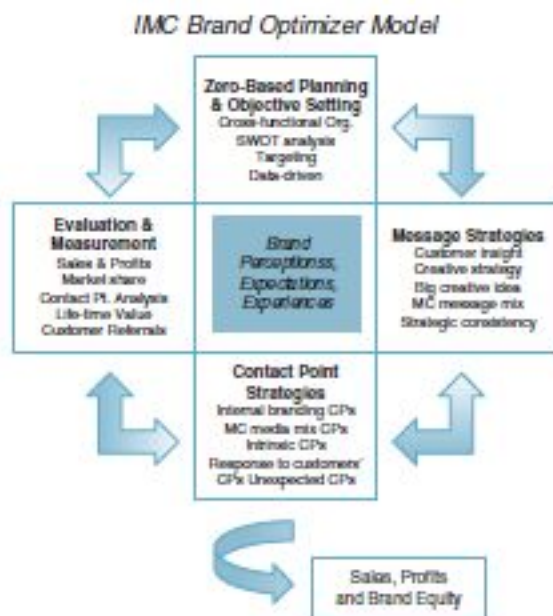


FIGURE 19.2

IMC Brand Optimizer Model

This IMC model created by Professor Tom Duncan illustrates that evaluation is a circular process—IMC plans start by gathering information and move through the various steps in the planning process to come back to the last step in the process, which is again gathering information. This time, however, information is gathered to determine what works and what doesn't. That information feeds back into the process and the organization learns from the results.

Principle

Good evaluation plans, as well as effective professional work, are guided by a model of how people respond to advertising.

A MATTER OF PRINCIPLE

Completing the Cycle

Mark Stuhlfaut, Assistant Professor, University of Kentucky



So you've analyzed the market up and down, honed a strategic position, spent months creating the most attractive image, produced hard-hitting materials, placed ads in all the right media, stretched the budget, made sure everyone in your marketing chain is on message, and launched a tightly integrated communication campaign. . . . It's time to sit back and enjoy the afterglow . . . right?

Not quite. Your job isn't finished until you've properly evaluated the results. Why? There are three very good reasons: (1) you need to find out what worked in the campaign, what didn't, and what could have worked better to solve any problems now; (2) a comprehensive evaluation provides valuable information that will serve as input for the next planning cycle; and (3) managers of marketing communication need to responsibly demonstrate the effectiveness of their efforts to clients and corporate management—you owe it to them to prove that their investment of resources in your programs and their trust in you were worth it.

Where do you start? A good beginning is to go back to the campaign's goals to see if they were met, which brings back the importance of having clear, measurable, and attainable objective statements.

What standards should you employ to know if you've succeeded? Sales data? They're one indicator, but too

many intervening factors make tying marketing communication to sales figures difficult and not very meaningful. Therefore, other measures, such as the levels of awareness, comprehension, importance, and brand preference, are more useful for communication managers to determine whether the campaign was effective.

A thorough review takes more than a few quick surveys. Sure, you'll want to conduct quantitative and qualitative research to get feedback from consumers or customers. But you should also contact all the key stakeholders in the market—such as company sales personnel, distributors, dealers, editors, broadcasters, consultants, and other friendly third parties (FTPs)—to see if their communication needs were satisfied by the campaign. Include these important people in your evaluation, and you'll not only gain helpful information, you'll build strong relationships for the future.

The best evaluation techniques aren't something you add on to a campaign; they're something you build in to every phase of the process. Assess alternative positions early in the campaign's development. Compare different concepts in the rough layout stage. The earlier you test, the cheaper it is, and the better chance you have to get it right.

It's easy to say you don't have the time or the money to evaluate the campaign's elements. But consider the cost of not getting it right. The effort made to evaluate the effectiveness of the campaign before and after launch will pay off in the long run.

The Vail Resorts' Colorado Pass Club launch is an example of how a model of effects can be developed for a specific campaign and used to drive not only the planning of the effort, but also the evaluation of its effectiveness. One of the goals to a successful ski season, according to Vail's Jennifer Wolfe-Kimbell, is to lock in revenue as early as possible. This can happen in multiple ways. For Colorado ski resorts, two groups of skiers bring in the bucks: in-state season pass holders and out-of-state visitors or "destination skiers." In-state skiers purchase season passes prior to the ski season and provide a surge of revenue going into the season. Destination skiers provide incremental revenue throughout the season by booking vacations and purchasing the lion's share of daily lift tickets, lodging and dining sales, ski school lessons, and rentals.

WHAT IS MESSAGE EVALUATION?

Copytesting is a general term that describes various kinds of research used at different stages in the advertising development process—before, during, and after an ad or campaign has run. Before we talk about specific types of studies, consider first the various copytesting services and the factors they feel are important to evaluate.

Table 19.1 Effectiveness Research Questions

Effect	Research Questions
Perception	
Awareness/noticed	Which ads do you remember seeing? Which ads were noted?
Attention	What caught your attention? Did the ad stand out among the other ads and content around it? What stood out in the ad?
Recognition (aided)	Have you seen this ad/campaign? Sort elements into piles of remember/don't remember.
Relevance	How important is the product message to you? Does it speak to your interests and aspirations?
Emotional/affective	
Liking/disliking	What emotions did the ad stimulate? How did it make you feel? Do you like this brand? This story? The characters (and other ad elements)? What did you like or dislike about the brand? The ad?
Desire	Do you want this product or brand?
Cognition	
Interest	Did you read/watch most of it? How much? Did it engage your interest or curiosity? Where did your interest shift away from the ad?
Comprehension/ confusion	What thoughts came to you? Do you understand how it works? Is there anything in the ad you don't understand? Do the claims/product attributes/benefits make sense? Do you have a need for this brand or can it fulfill a need for you?
Recall (unaided)	What happened in the commercial? What is the main message? What is the point of the ad?
Brand recall/linkage	What brand is being advertised in this ad? [In open-ended responses, was the brand named?]
Differentiation	What's the difference between Brand X and Y?
Association	
Personality/image	When you think of this brand, what (products, qualities, attributes, people, lifestyles, etc.) do you connect with it? Do you link this brand to positive experiences? What is the personality of the brand? Of whom does it remind you? Do you like this person/brand personality? What is the brand image? What does it symbolize or stand for?
Self-identification	Can you see yourself or your friends using this brand? Do you connect personally with the brand image?
Persuasion	
Intention	Do you want to try or buy this product/brand? Would you put it on your shopping list?
Argument/ counterargument	What are your reasons for buying it? Or for not buying it—or its competing brand(s)? How does it compare to competitors' brand(s)? Did you argue back to the ad?
Believability/ conviction	Do you believe the reasons, claims, or proof statements? Are you convinced the message is true? The brand is best?
Trust	Do you have confidence in the brand?
Behavior	
	How many people buy, try, call, send, click, visit, attend, inquire, volunteer, donate, advocate, or whatever the desired action? What is the rate of change?

Copytesting

Copytesting companies have different specialties that focus on different effectiveness dimensions. The most successful of these companies have conducted enough tests that they have developed norms for common product categories. In other words, after they pretest an ad, they can compare its score with scores from comparable ads. Norms allow the advertiser to tell whether a particular advertisement is above or below the average for the brand or its product category. Without norms the advertiser would not know whether a score of 23, for example, is good or bad.

Most of these companies have also developed diagnostic methods that identify strong and weak points of the ad. Here is a list of some of the more prominent companies and the types of tests they provide:

- *Ameritest*: brand linkage, attention, motivation, communication, flow of attention and emotion through the commercial
- *COMSCORE ARSgroup*: persuasion, brand/ad recall, communication
- *Ipsos ASI*: recall, attention, brand linkage, persuasion (brand switch, purchase probability), communication
- *Mapes & Ross*: brand preference change, ad/brand recall, idea communication, key message delivery, like/dislike, believability, comprehension, desire to take action, attribute communication
- *Millward Brown*: branding, enjoyment, involvement, understanding, ad flow, brand integration, feelings about ad, main stand-out idea, like/dislike, impressions, persuasion, new news, believability, relevance
- *RoperASW*: overall reaction, strengths and weaknesses, understanding, clutter busting, attention, main message, relevance, appeal, persuasiveness, motivate trial, purchase intent.

Now let's consider how the services of such companies and the research they conduct can be used at various points in the development and evaluation of a marketing communication effort. This outline is based on the Facets Model of Effects that we first introduced in Chapter 4.

Message Development Research

Deciding what facts to convey in marketing communication is never easy. Research is needed to develop and test alternative message strategies. Planners conduct research with members of the target audience to develop the message strategy and test the relative effectiveness of various selling premises—hard sell or soft sell, informational or emotional, and so forth. Insights into consumer motivations and purchasing decisions help solve the often-difficult puzzle of selecting the most relevant information and motivating promise as well as the emotional appeal that best engages the audience.

Concept Testing Advertising and other marketing communication messages usually incorporate a Big Idea, a creative concept that is attention getting and memorable. Research in **concept testing** compares the effectiveness of various message strategies and their creative ideas. This testing often relies on a **key concept card**, which is an artist's drawing of the visual idea with a sentence underneath that captures the essence of the idea. A researcher may use a pack of three, five, or more idea cards or rough layouts to elicit consumer responses in a mall or through discussions in a focus group.

An example of effective concept testing that was used in a campaign's development is Volvo's GLBT campaign, which was the first campaign of its kind to receive an Advertising Research Foundation (ARF) David Ogilvy award. Witeck-Combs Communications in collaboration with Prime Access, Inc., developed a set of concepts and rough ad executions representing a range of GLBT-specific imagery and copy. Ford and Volvo managers selected their top three choices from the group of proposed ad ideas. The messages were tested to assess the constructs of branding, communication, and persuasion, including the concept of consumer connection, or whether the ad makes consumers feel closer to the brand. These constructs were measured in four ways—cognitive, behavior, emotional, and aspirational dimensions.

Pretesting Another type of evaluative research, called **pretesting**, helps marketers make final go/no-go decisions about finished or nearly finished ads. Pretesting differs from concept testing

or message strategy research, which reveals the strengths and weaknesses of different versions of a concept or approach as marketers develop them. Pretesting assesses the strength of the finished message and predicts how well it will perform.

In terms of print advertisements, the ideas to be tested are often full mockups of the ad. In television, advertisers can test the commercials as storyboard ideas or photoboards (still photos arranged as a storyboard), but more often commercials are in the form of animatics (drawings or still photos shot on videotape synchronized with a rough version of the audio track). Advertisers can show these versions of the commercial to individuals, to focus groups, or in a theater testing situation. They follow the viewing of the advertisement with a survey, a more open-ended interview, or a set of questions for a group discussion.

Diagnostics Many advertisers and agencies are moving away from copytesting methods that rely on single scores to evaluate an ad and turning to methods that are more focused on diagnosing strengths and weaknesses. The reason is that they believe an advertisement is too complex to be reduced to one factor and one simple score. Instead they are using research methods designed to diagnose the strengths and weaknesses of their advertising ideas to improve the work while it is still in development or to learn more from the ad to improve subsequent advertisements.

In theater tests for TV commercials, for example, respondents may have a black-box device and can press a button to record different types of responses—indicating what they liked or didn't like or how long they paid attention by letting up on the button when their attention shifts.

Moment-by-moment tests of commercials provide an analysis of the impact of the internal logic of the commercial. The Ameritest company, whose work is described in the *A Matter of Practice*, looks at two dimensions of moment-by-moment tests: a *Flow of Attention* measure and a *Flow of Emotion* measure. The procedure involves showing a clutter reel, a group of commercials that includes the test commercial, competitors' commercials, and other ads, and conducting interviews afterward. The Ameritest method uses a picture sort to diagnose the viewer's attention to and emotional engagement with different elements in the commercial. The viewers receive a deck of key frames from the commercial and sort them into images they remember seeing and ones they don't remember. Then they sort them into five categories from strong positive to strong negative. The researchers tabulate the sets to depict the flow of impact for both attention and emotion. In particular, they want to identify and analyze key moments in the commercial such as the solution to a problem or the introduction of the brand and analyze them in terms of viewers' attention and emotion.⁶ The *A Matter of Practice* box takes you through this analysis.

Principle

Advertising effects are too complicated to be reduced to a single score.

During Execution: Concurrent Testing

Concurrent testing takes place while the advertising is running. There are three primary techniques: coincidental surveys, attitude tests, and tracking studies. The first two assess communication responses; tracking studies evaluate actual behavior.

The coincidental survey technique is most often used with broadcast media. Random calls are made to individuals in the target market. By discovering what stations or shows people are tuned to, the researcher can determine whether the target audience has seen/heard the ad and, if so, what information or meaning the audience members now have of the brand. This technique can be useful in identifying basic problems. For example, several years ago Pepsi discovered that the use of Madonna as a spokesperson was a terrible mistake.

In Chapter 4 we discussed the relationship between an attitude—a favorable or unfavorable disposition toward a person, brand, idea, or situation—and consumer behavior. This relationship is the basis of *attitude tests*. Researchers survey individuals who were exposed to the ad, asking questions about the spokesperson, the tone of the ad, its wording, and so forth. Results that show strong negative attitude scores may prompt the advertiser to pull an ad immediately. A favorable attitude indicates that people are more likely to purchase a brand than if they have an unfavorable attitude.

Tracking Studies Studies that periodically (generally every three or six months) measure top-of-mind brand awareness (first brand mentioned) are called *tracking studies*. These studies can

A MATTER OF PRACTICE

Finding Moments of Truth

Charles E. Young, President, Ameritest



The most powerful search engine of all is the human eye, which scans advertising film, television commercials, and Web videos, continuously deciding on an unconscious level whether the visual information streaming toward it is important enough to let into consciousness. Because our conscious minds have limited bandwidth or work-space, much of the imagery that advertisers are trying to communicate to consumers is ignored or deleted by our preconscious eye-brain filters as so much visual spam.

Ameritest's Picture Sorts® is a set of nonverbal research tools that have been developed for the internet to survey the right-brained scanning and sorting processes involved in visual communication. These tools make use of the power of still photographs to capture an instant of time and store our fleeting emotions.

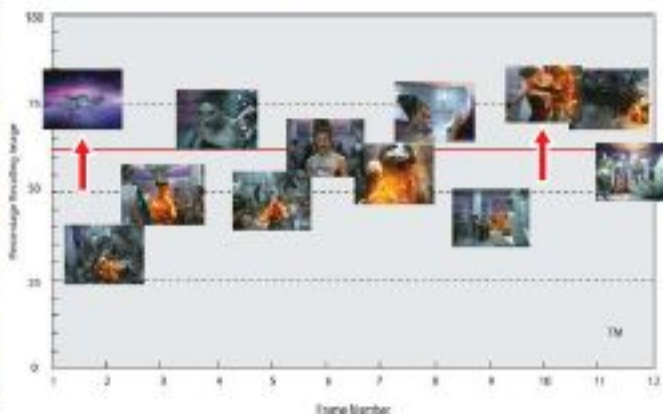
By sorting a randomized deck of pictures taken from the ad itself—which is like the visual vocabulary of the film—the ad researcher can reconstruct consumers' moment-by-moment attention and emotional response to an ad they just saw. Three different sorting exercises enable the advertiser to perform the equivalent of putting on 3-D glasses to see advertising through the eyes of its target audience.

The Flow of Attention® graph, the first of three measurement dimensions, is like a visual spell-checker that the researcher can use to analyze whether or not a piece of advertising film or Web video has been put together well according to principles of proper film syntax or good visual grammar. The Flow of Attention graph reveals the hidden structure of audience attention to moving pictures, which, like music, follows a rhythmic beat of cognitive processing. The beat, or focal points of attention, is where the most important information in an ad, like the brand identity, is conveyed.

From the emotional hook at the beginning, to the turning points in a story, to the surprise ending of a funny commercial, engaging the emotions of con-

sumers is essential to motivating them. The Flow of Emotion® graph measures not only the volume of emotions pumping through ad film but also reveals which of four archetypal dramatic structures is being used in the creative design. Knowing this structure tells the advertiser when the timing is exactly right to first introduce the brand in the ad, which might be at the beginning, or somewhere in the middle, or not until the end of a commercial.

The Flow of Meaning® tool shows the researcher where key communication points or brand values are being cued visually. Meaning is created when thought and emotion come together, in a few memorable and emotionally charged moments in a commercial when memories are being created. Because there are three distinct memory systems in the mind, branding moments come in three flavors: (1) images that convey concepts or rational ideas go into our knowledge, or semantic, memory system; (2) images that evoke emotions go into our emotion, or episodic, memory system; and (3) images that rehearse or mirror the behavior the advertiser is trying to influence go into our action, or procedural, memory system (where memories of how you ride a bike or play a violin are stored). Taken together, this learn-feel-do imaging process is how the long-term work of advertising is performed, building a brand's image.



This diagram demonstrates the flow of attention across a commercial for Unilever's Thermafilk. The selected still frames represent those places in the commercial where the attention is high (above the red line) or low (below the red line). Note the highest attention points, which are indicated with the red arrows. In this commercial, that tight shot of the face is the "moment of truth," the most highly attended to frame in this execution.

also measure unaided and aided awareness of an ad and trial and repeat purchases. They sometimes ask the same questions of competing brands.

Tracking studies show how a marketer's brand is performing over time, especially after changes are made in the marketing communication. Because several different measures are taken, findings can indicate if there is an attention-getting problem, a recall problem (saw message but can't remember what it said), a trial problem (remembered the advertising but not moved to try the brand), or a repeat problem (tried the brand but have not repurchased for some reason).

Brand tracking is an approach that tracks the performance of the brand rather than or in addition to the advertising. The assumption is that with fragmented media and an abundance of high-quality but similar products, it is more important to track the brand because it reflects the quality of the customer's brand relationship. That's at the heart of the Colorado Pass Club campaign. Instead of focusing on attributes and claims about a product, this research identifies how customers are involved with the brand and whether they are more favorably disposed toward it than toward other brands. As Peppers and Rogers point out, you also need a centralized customer database that allows you to track customer activity across channels.⁷

This analysis as depicted in Figure 19.3 is also based on the factors identified in the Facets Model of Effects.

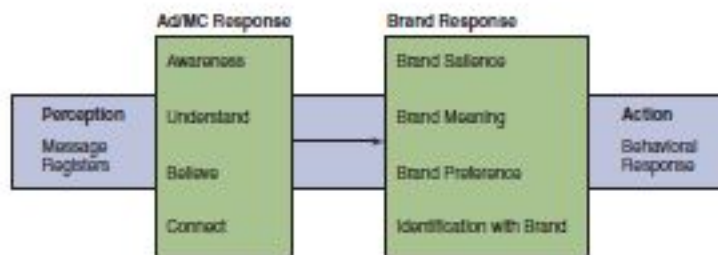


FIGURE 19.3
Tracking Brand Response

Brand tracking studies generally use sets of brand-related questions, but Bain & Company, a highly respected consulting firm, is promoting the use of a single-question test to track attitudes to a brand. Called *net promoter*, it tracks the response to a simple question: "How likely is it that you would recommend us to a friend or colleague?"⁸ Another single measure is **brand penetration**, which is the number of customers who purchase the brand relative to the total population in the market.⁹

Although these single measures are useful as a broad indication of brand effectiveness, the other questions found in brand tracking studies, such as those about awareness and credibility, for example, are particularly important in an IMC program because they provide diagnostic information about areas of marketing communication that might need to be changed or refined.

Because spending information enters the analysis, much of the focus of tracking studies is on the target market, the selection of media vehicles, the schedule, the marketing communication mix, and the media mix. Account planners use several methods to collect tracking data, such as the attitude tests discussed earlier and wave analysis, consumer diaries, and pantry checks:

- **Wave Analysis** Wave analysis looks at a series of interviews during a campaign. The tracking begins with a set of questions asked of a random sample of consumers on a predetermined date. After the person is qualified as hearing or seeing the ad, the researcher asks a series of follow-up questions. The answers serve as a benchmark and allow adjustments in the message content, media choice, and timing. Perhaps two months later, the researcher makes another series of random calls and asks the same questions. The second wave is compared with the first until management is satisfied with the ad's market penetration and impact.
- **Consumer Diaries** Sometimes advertisers ask a group of representative consumers to keep a diary during a campaign. The consumers record activities such as brands purchased, brands used for various activities, brand switches, media usage, exposure to competitive promotions,

and use of coupons. The advertiser can then review these consumer diaries and determine factors such as whether the message is reaching the right target audience and whether the audience is responding to the message as intended. One common unfavorable finding from consumer diaries is that no attitude or behavioral change occurred because of exposure to the campaign.

- **Pantry Checks** The pantry check provides much of the same information as the diary method but requires little from the consumer. A researcher goes to homes in the target market and asks what brands or products they have purchased or used recently. In one variation of this procedure, the researcher counts the products or brands currently stocked by the consumer. The consumer may also be asked to keep empty packages, which the researcher then collects and tallies.

Test Markets A test market might serve to evaluate product variations, as well as elements of a campaign or a media mix. These are generally conducted in two or more markets with the same number of similar markets chosen to act as controls. In the control markets the researcher can either (1) run no advertising or (2) continue to run the old ad. The new advertising (or product variation) is used in the test cities. Before, during, and after running the advertising, marketers compare sales results in the test cities. Some cities, such as Buffalo, Indianapolis, and San Antonio, are considered excellent test markets because their demographic and socioeconomic profiles are representative of either the United States or a particular target market. Furthermore, they are relatively isolated as media markets so the advertising impact is less likely to be affected by what is happening in other markets.

The possibilities for isolating variables in test markets are almost limitless. Researchers can increase the frequency of advertising or try a different media schedule, for example. They can see whether an ad emphasizing brand convenience will stimulate sales to two-career families. They can try an ad that plays up the brand's fiber or vitamin content or compare the effectiveness of a two-for-one promotion versus a cents-off coupon.

Post-Testing: After-Execution Research

Another evaluation method is post-testing, which occurs after the marketing communication has been used to determine if it met its objectives. The most common evaluative research techniques account planners and advertising researchers use include memory tests, persuasion tests, direct-response counts, frame-by-frame tests, in-market tests such as test markets, and brand tracking.

Some of these more traditional post-testing methodologies were discussed in Chapter 6, however, there is another type called *heuristics* that is relevant for certain types of post-testing evaluation. Harley Manning, vice president for customer experience at Forrester Research, describes heuristics as a type of expert review that Forrester uses to evaluate the performance of websites in terms of brand attributes (information provided) and user descriptions and goals as measured against "a valid set of rules (heuristics) that identify known types of user experience flaws." Manning explains that Forrester uses that type of expert-review research to gauge how well the sites served consumers (keep the brand promise).

The idea is that advertising—and other types of marketing communication messages—vary in the way they affect people, so the measurements also must be different.

Breakthrough Attention Most advertising is evaluated in terms of its ability to get and keep attention. This is a simple concept but difficult to measure. Similar to pretesting, some researchers use instantaneous tracking in a theater setting where viewers with a keypad indicate what they watch and don't watch—or rather what interests them and doesn't interest them. The Ameritest methodology asks respondents who have just watched a collection of ads to indicate which ones interested them.¹⁰ Another firm's method asks viewers to rate how enjoyable the ad is, which confounds the attention score with a liking response. Other methodologies use noted scores for print ads or ask viewers to recognize the concept of a commercial and attach it to a brand.

Engagement Tests Interest in a commercial is determined by concentration and excitement. Eye-tracking research is a mechanical technique that traces saccadic eye movement, the path the eye

takes in scanning an ad. Because scanning involves a lot of stops and starts and revisits, this complex map of how a visual is scanned identifies what first caught the viewer's attention—the visual point of entry—and where it moved to next. It also records those elements the eye kept revisiting either to focus on because they were appealing visuals or because they needed more study. It also shows what elements didn't get noticed at all. The e-Motion system from the New Jersey-based firm PreTesting Group identifies eye fixations but also the amount of vibrations in the retina. The vibration is very minute, but it does indicate a level of excitement—the more the retina vibrates, the more interested the viewer.¹¹

Understanding and Comprehension Tests These tests are used as part of the diagnostic effort, but it is particularly important in post-testing to determine if consumers understood the message. Another comprehension problem can emerge from creative strategies that are engaging, but ask consumers to puzzle out the meaning. Ambiguity is sometimes used particularly because it is engaging, but it can also be a risky strategy, one that benefits from testing. An example is a billboard for a British beer brand John West, which is given a dramatic touch by making the rings on top of the can look like rings of rippling water cast from a fishing bobber.



A beautifully done image for John West, a British beer, uses a simple photograph of the top of a can. It takes on new meaning when the can's rings are associated with the rings in water from a fishing bobber. The question is do people get the association and understand the larger meaning of "Relax with a Can of John West."

Memory Tests Memory tests are based on the assumption that an advertisement leaves a mental residue with the person who has been exposed to it; in other words, the audience has learned something. One way to measure an advertisement's effectiveness, then, is to contact consumers who saw the ad and find out what they remember. Memory tests fall into two major groups: recognition tests and recall tests.

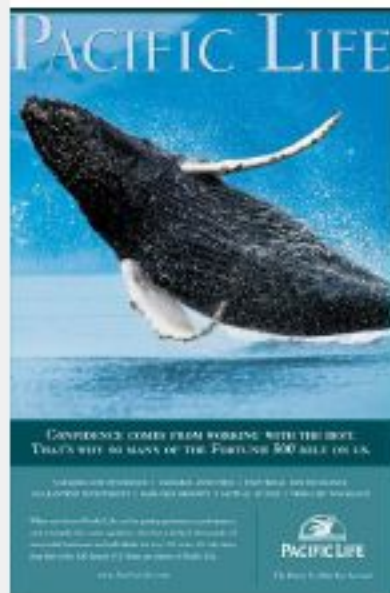
One way to measure memory is to show the advertisement to people and ask them whether they remember having seen it before. This kind of test is called a **recognition test**. In a **recall test**, respondents who have read the magazine are asked to report what they remember from the ad about the brand. In **aided recall tests**, the interviewer may go through a deck of cards containing brand names. If the respondent says, "Yes, I remember seeing an advertisement for that brand," the interviewer asks the interviewees to describe everything they can remember about the ad. Obviously recall is a more rigorous test than a recognition test.

Similarly, a TV commercial is run on network television within a regular prime-time program. The next evening, interviewers make thousands of random phone calls until they have contacted about 200 people who were watching the program at the exact time the commercial appeared. The interviewer then asks a series of questions, such as the following:

- Do you remember seeing a commercial for any charcoal briquettes?
- *If No* Do you remember seeing a commercial for Kingsford Charcoal briquettes? (memory prompt)
- *If Yes to Either of the Above* What did the commercial say about the product? What did the commercial show? What did the commercial look like? What ideas were brought out?

The first type of question is called an **unaided recall question** because the particular brand is not mentioned. The second question is an example of **aided recall**, in which the specific brand name is mentioned. The answers to the third set of questions are written down verbatim. The test requires that the respondent link a specific brand name, or at least a specific product category, to a specific commercial.

Brand Linkage Tests Testing for brand linkage is basically a test of memorability, however, the important aspect of this measure is not whether the message is remembered, but whether it



Pacific Life uses an image of a leaping whale to reflect its image of a confident insurance company that excels in its market.

associates the brand with the memory. If the commercial fails to establish a tight connection between the brand name and the selling message, the commercial will not get a high recall score. This is also called a test of “brand linkage.” The long-running Pacific Life campaign with its image of a leaping humpback whale is a good example of a visual that serves as a strong brand reminder and associates the brand with strength and performance.

Emotion Tests Advertising is just beginning to move into this area of research. There are several ways to get at the emotional response. Ameritest uses photo sorts with still frames from commercials to identify the positive and negative moments that touch people’s emotions.¹² Other researchers are wiring up viewers of television ads to monitor brain activity using a functional magnetic resonance imaging (fMRI) machine. The fMRI images identified ads from the 2007 Super Bowl that produced anxiety, fear, anger, and insecurity, as well as positive feelings.¹³

Likability Tests A study by the Advertising Research Foundation (ARF) compared a variety of different copytesting methods to see if any of them were better able to predict sales impact. Surprisingly, it wasn’t awareness, recall, communication, or persuasion measures that won out but rather **likability tests**. Likability, however, is not easy to measure because it’s difficult to know if the consumer likes the ad, the brand, or some other factor, such as the person giving the test. A number of the copytesting companies offer a likability score, but they suggest it needs to be interpreted relative to other consumer responses. Questions that try to evaluate likability investigate factors such as these: personally relevant, important to me, stimulates interest or curiosity, creates warm feelings, enjoyable, entertaining, and fun.

Persuasion Tests Another evaluative research technique is a **persuasion test**, or attitude change test. The basic format is to ask consumers how likely they are to buy a specific brand. Next they are exposed to an advertisement for that brand, usually as part of a collection of brands. After exposure, researchers again ask them what they intend to purchase. The researcher analyzes the results to determine whether *intention to buy* has increased as a result of exposure to the advertisement. This test is sometimes referred to as an **intend-to-buy** or **motivation test**. The validity of a persuasion test depends in part on whether participants in the experiment represent a good sample of the prospects the advertiser is trying to reach. A dog food advertiser, for example, would not be interested in responses from people who do not own dogs.

Inquiry Tests A form of action response, **inquiry tests** measure the number of responses to an advertisement. The response can be a call to a toll-free number, an e-mail or website visit, a coupon return, a visit to a dealer, an entry in a contest, or a call to a salesperson. Inquiry tests are the primary measurement tool for direct-response communication, but they also are used to evaluate advertisements and sales promotions when the inquiry is built into the message design. Inquiry tests also are used to evaluate the effectiveness of alternative advertisements using a split-run technique in magazines, where there are two versions of the magazine printed, one with ad A and the other with ad B. The ad (or direct-mail piece) that pulls the most responses is deemed to be the most effective.

Scanner Research

Many retail outlets, especially drug, discount, and food stores, use electronic scanners to tally up purchases and collect consumer buying information. When you shop at your local Safeway, for example, each product you buy has an electronic bar code that conveys the brand name, the product code, and its price. If you are a member of Safeway’s frequent buyer program and have a membership card that entitles you to special promotional offers, the store can track your purchases.

Scanner research is also used to see what type of sales spikes are created when certain ads and promotions are used in a given market. Both the chain and the manufacturers of the brands are interested in this data. The regional Safeway system may decide to establish a consumer panel so it can track sales among various consumer groups. In scanner research, you would be asked to join a panel, which might contain hundreds of other customers. You would complete a fairly extensive questionnaire and be assigned an ID number. You might receive a premium or a discount on purchases for your participation. Each time you make a purchase, you also submit your ID number. Therefore, if Safeway runs a two-page newspaper ad, it can track actual sales to determine to what extent the ad worked. Various manufacturers who sell products to Safeway can do the same kind of testing. The panel questionnaire also contains a list of media that each member reported using, so media performance can also be evaluated.

Using scanner data and the cooperation of local cable networks, researchers are closer to showing a causal relationship between advertising/promotion and sales because of **single-source research**. Single-source research companies, such as A. C. Nielsen, arrange to have test commercials (and sometimes test newspaper ads) delivered to a select group of households (HHs) within a market, comparing it to a control group of HHs. The purchasing behavior of each group of HHs is collected by scanners in local stores. Because advertising is the only manipulated variable, the method permits a fairly clear reading of cause and effect. Data collected in this way are known as *single-source data* because advertising and brand purchasing data come from the same HH source.

Syracuse University Professor John Philip Jones, who spent many years at J. Walter Thompson (JWT), has used single-source data from JWT combined with Nielsen TV viewing data to prove that advertising can cause an immediate impact on sales. His research has found that the strongest campaigns can triple sales, while the weakest campaigns can actually cause sales to fall by more than 50 percent.¹⁴

Although fairly expensive, single-source research can produce dependable results. Advertisements are received under natural conditions in the home, and the resulting purchases are actual purchases made by consumers. One drawback, besides cost, is the three to six months required to set up and run this test. Critics also say that single-source research is better for short-term immediate sales effects and doesn't capture very well other brand-building effects.



Scanner research reads the information from a shopper's identification card and records that along with product information. Many retail outlets use electronic scanners to track sales among various consumer groups.

MEDIA EVALUATION

Advertising has little chance to be effective if no one sees it. Analyzing the effectiveness of the media plan is another important part of evaluation. Did the plan actually achieve reach and frequency objectives? Did the newspaper and magazine placements run in the positions expected and produce the intended GRP and CPM levels? In other words, did the advertisers get what they paid for?

A classic case of the power of media, particularly the repetition of commercials, is the "Please Don't Squeeze the Charmin" campaign that began in 1964. Because it dominated the airwaves for 20 years, it was hugely visible and built incredible brand awareness—awareness that led Charmin to dominate its category. It and its brand character, the comical grocer Mr. Whipple, was not only one of the best remembered campaigns, but also one of the most parodied campaigns ever made because the Whipple character was so goofy. Yet, it made Charmin one of the most successful brands in the history of Procter & Gamble. So what do you think: Does it make sense to use an irritating campaign with a lot of repetition in order to build brand awareness and memory? Could this strategy work in our time? What would an IMC planner say about this strategy?

Evaluating Audience Exposure

For major campaigns, agencies do post-buy analyses, which involve checking the media plan against the performance of each media vehicle. The critical question is whether the reach and frequency objectives were obtained.

Verifying the audience measurement estimates is a challenge. Media planners are working sometimes with millions of dollars, and they can't afford to get it wrong. We discussed how various media channels measure their audiences. For print, services such as the Audit Bureau of Circulations (ABC), Experian Simmons (formerly known as SMRII), and Mediamark (MRI+) provide data. Likewise for broadcast, Arbitron, RADAR, and A. C. Nielsen provide audience monitoring. Initially media planners use these estimates to develop a media plan, and media buyers use them later to verify the accumulated impact of the media buy after the campaign has run.

Media planning oversight is usually handled in-house by the media buyer, but it can also be contracted by the advertiser to independent companies who specialize in conducting media audits of the agency's media planning and buying operations. Nissan and Procter & Gamble are examples of companies that have hired outside media-auditing firms to confirm the execution of their media plans.

As the job gets more complex, media planners are being asked to prove the wisdom of their recommendations in an area where the data they use are sometimes suspect or unreliable, particularly if there are problems with the media measurement companies' formulas and reporting systems. Nielsen, for example, has been subject to much criticism for its television ratings.

Another issue is the impact of video recorders and DVRs. What do viewers see and remember as they skip or fast forward through commercials? Tests have shown that viewers do remember some of the spots, or at least the brands. The easiest to recall were ones that had been seen before. Other characteristics of successful ads have the brand's logo in the middle of the screen and leave it on the screen for more seconds than in a normal TV commercial. It also helps if the action isn't too fast with many screen changes. Simple commercials are easier to see and remember when zipping.¹⁵

To better understand the problems in media evaluation, let's look at two areas where media performance is hard to estimate: out-of-home media and new media, including the Internet.

Out-of-Home Media As you would expect, accurately measuring the mobile audience for outdoor advertising is challenging. Traffic counts can be gathered, but the problem is that traffic does not equal exposure. Just because a car drives by a board doesn't mean that the driver and/or passengers see it, particularly since some outdoor boards are more attention getting than others, as the "road rage" board illustrates.

New Media Similarly, the measures of effectiveness used to evaluate offline campaigns don't seem to transfer well to the online world. Is the website visitor or banner ad viewer similar to a member of the print or broadcast audience? The online industry hasn't been able to establish on-

line equivalencies for GRPs and CPMs and is still trying to define what makes an effective Internet ad and to develop a system that accurately measures online effectiveness. At the heart of the problem is the question of what exactly is to be measured and how that equates to other media: readers, viewers, visitors, hits, click-throughs, or minutes spent with a site? Web-analytic firms are developing more sophisticated tracking programs to measure user activity in terms of Web pages visited, time spent, and click-throughs.

Alternative media, such as word of mouth, social media, and guerilla marketing campaigns, are even harder to measure and media planners search for reliable indicators of exposure numbers and buzz from these new sources that equate to the performance of traditional mea-



This outdoor board from the United Kingdom attracted attention because of its interesting visual but also because of its challenging idea. Research based on traffic counts find it difficult to account for the emotional impact of messages like these.

sured media. Research company Millward Brown has designed a metric for online word of mouth (WoM) to track and analyze sentiments expressed on social networks, blogs, and chat rooms.¹⁶ Procter & Gamble has created TREMOR, which develops buzz campaigns, but is also used to design analysis techniques to track measurable business results for WoM campaigns.

An example of using new media analytical tools to assess performance comes from the Taco Bell “Drive-Thru Diet” campaign, which promoted the fast-food chain’s “Fresco” menu. Although the campaign wasn’t presented as a weight-loss program, it generated a significant amount of buzz questioning the strategy. According to Zeta Buzz, a firm that tracks postings on blogs, message boards, and social media, Taco Bell’s buzz rating dropped six points—the volume of postings increased but the tone was much more negative, which indicated the strategy was backfiring.¹⁷

Another interesting experiment was conducted by Boston University college students and the Mullen ad agency using Twitter to get near-instantaneous feedback on how viewers were reacting to Super Bowl ads. The project was designed “to use a new medium to comment on an old medium,” according to Mullen’s chief creative officer.¹⁸

ROI and Media Efficiency

Advertisers continue to improve how they measure advertising ROI (return on investment, which means the costs of creating and running the advertisement versus the revenue it generates). Another way to look at it is the cost-to-sales ratio.

Since the dollar impact for advertising—and public relations, especially—is difficult to measure, ROI is hard to calculate. The campaigns must be carefully designed not only to increase sales, but also to ensure that advertisers can isolate the impact of the message and verify that the advertising caused the increase in sales. ROI is easier to calculate for direct marketing (because there are fewer variables between the message and the sales) and for sales promotion (because there is an immediate response, which is easier to link to the message).

One question related to ROI is this: How much spending is too much? That is, how do you determine whether you are overadvertising or underadvertising? That’s one of the key reasons to use test marketing. If a campaign is launched in several different, but matched, cities at different levels of media activity, a comparison of the campaign results (sales or other kinds of trackable responses) can determine the appropriate level and type of media spending.

Wearout The point where the advertising, because it has been seen multiple times, no longer stimulates much of a response is called **wearout**. This is also the point where recall stabilizes or declines and irritation levels increase because people are tired of hearing or seeing the same ad.

Wearout is a combination of creative impact and media buying. The more intrusive or the less interesting the creative technique, the higher the level of irritation. It’s like a joke: you may pay attention the first couple of times you hear it, but then it gets “old.” Other types of advertising are less prone to wearout. Good jingles, for example, can be repeated almost endlessly. The more people like to hum along, the less likely there will be a wearout problem.

The issue is how long and how much is needed to create the necessary impact. John Philip Jones, who has done extensive research into effectiveness issues, argues that research supports higher continuity. He recommends spreading the budget out and adding weeks to the media schedule as much as possible.¹⁹ Another way to evaluate such decisions is through media optimization models.

Media Optimization One of the biggest challenges in media planning is media efficiency—getting the most for the money invested. You may remember from Chapter 14 that media planners operate with computer models of **media optimization**—or the optimum media performance—that are used in making decisions about media selection, scheduling, and weights (amount of budget). Models are always theoretical, so one result of postevaluation is that the actual performance of a plan can be compared with the results projected by the media planner’s model. The goal of media optimization is to optimize the budget—to get the most impact possible with the least expenditure of money. That is the critical finding derived from the comparison of performance with projections. In addition to meeting the reach and frequency objectives, was the media plan efficient?

A new dimension to the efficiency question comes from a program by media company Starcom to measure the effectiveness of various media used by client Applebee's. The objective was to develop a model that compares the value of video ads, both on-air and online, across a number of NBC properties. The model will establish the relative value of the ads and also interaction across channels.²⁰

EVALUATING IMC TOOLS, CAMPAIGNS, AND PROGRAMS

Evaluation is the final and, in some respects, most important step in an advertising campaign because it determines whether the campaign effort was effective. A campaign evaluation will assess the campaign's performance in terms of its message and media objectives. It employs either internal performance data or results from an outside research organization that monitors brand performance, such as Gallup-Robinson, Decision Analyst Inc., and Millward Brown. The important thing to remember in planning an evaluation program is that evaluation has to be built into the marketing communication plan from the very beginning.

An IMC campaign employs various marcom tools, such as sales promotion, public relations, direct marketing, events and sponsorships, personal sales, packaging, point of purchase (PoP), and specialties. You'll find that most of these areas have their own metrics by which performance is measured.

Even though an IMC campaign is evaluated in terms of its overall impact on a brand, the pieces of the mix may also be evaluated to determine the effectiveness of their performance. Advertising may be the most visible; however, other marketing communication tools, such as sales promotion, are better at getting people to respond with an immediate purchase, and public relations is particularly strong at building credibility. Let's look briefly at these other marketing communication areas in terms of their evaluation.

Marcom Tools

As Table 19.2 illustrates, marketers can use the Facets Model of Effects to identify the objectives commonly associated with the various marketing communication areas as well as the types of measures used to evaluate performance. The idea is that certain marketing communication functions, such as public relations and sales promotion, do some things better than other areas. Therefore, in an integrated plan, we would use the best tool to accomplish the desired effect. In Table 19.2, the main effects are located in the first column, with a collection of surrogate measures identified in the second column (this list is not inclusive, it's just a sample). The last column lists the communication tool or tools that may be most appropriate for achieving the objective.

Principle

Advertising is particularly effective in accomplishing such objectives as creating exposure, awareness, and brand image and delivering brand reminders.



Evaluation doesn't just happen at the end of a campaign or after the ad is run. It has to be planned into the campaign from the very beginning as this planning meeting illustrates.

Advertising How do you describe effective advertising? You've been watching it for most of your life. One brand manager's answer to that question is presented in the *Practical Tips* at www.pearsonhighered.com/moriarty. An examination of Table 19.2 shows that advertising is particularly effective in accomplishing a number of objectives, such as creating exposure, awareness, and a brand's image. It is also good at providing brand reminders to the customer and encouraging repurchases. The preceding discussions about copytesting, message development research, concurrent testing, post-testing, and the evaluation of media cover the most common advertising methodologies. Many of these research methods are used by other marcom areas, but they all are derived initially from advertising.

Direct Response The objective of direct-response communication is to drive a transaction or generate

Table 19.2 Message Effectiveness Factors

Key Message Effects	Surrogate Measures	Communication Tools
Perception	Exposure	Adv media; PR, PoP
	Attention	Adv; sales promo (SP), packaging; PoP
	Interest	Adv; SP; PR, direct; PoP
	Relevance	Adv; PR; direct; PoP
	Recognition	Adv; PR, packaging; PoP, specialties
Emotional/Affective	Emotions and liking	Adv; SP, Packaging; PoP
	Appeals	Adv; PR; sales; events/sponsorships
	Resonate	Adv; PR; events/sponsorships
Cognition	Understanding	Adv; PR; sales; direct
	Recall	Adv; SP; PR, PoP, specialties Adv; PR; packaging
Association	Brand Image	Adv; PR; events/sponsorships
Persuasion	Attitudes	Adv; PR; direct
	Preference/Intention	Adv; PR; sales; SP
	Credibility	PR
	Conviction	PR; sales; direct
	Motivation	Adv; PR; sales; SP
Behavior	Trial	SP; sales; direct, PoP
	Purchase	SP; sales; direct
	Repeat purchase	Adv; SP; sales; direct, specialties

some other type of immediate behavioral response, such as a donation or visit to a dealer. What makes this marketing communication tool so attractive to marketers is that response is so easily measurable. Some advertisements request direct response via a toll-free number, a mail-in coupon, a website or e-mail address, or an offer embedded in the body copy. Instead of depending on consumers' memories, measures of a message's persuasive abilities, or some other indirect indication of effectiveness, the advertiser simply counts the number of viewers or readers who request more information or buy the product.

In some ways, direct response mechanisms are the easiest marketing communication tool to evaluate both in terms of message effectiveness and in terms of ROI efficiency—the sales-to-cost ratio. The efficiency of a direct-response offer is measured in terms of responses per thousand (RPM). To calculate the RPM, use the following formula:

$$\frac{\text{Total responses}}{\text{Total mailed}} \times 1,000 = \text{RPM}$$

This calculation lets you compare the response rate of alternative mailings. For example if one mailing of 15,000 pulled 750 responses, then the RPM was 50 per 1,000. If a different mailing of 12,000 pulled 800 responses, then the RPM was 66 per 1,000, making it a more effective offer.

Sales Promotion Sales promotion programs for packaged goods and other products that use distribution channels need to evaluate both the impact of consumer (or end-user) direct promotions and promotions targeted at retailers and other channel members. You will recall the discussions of promotional allowances offered to retailers who agree to feature a brand in their ads or provide special in-store display and price discounts. These are measured by proof of performance such as

Principle

Direct-response communication drives action and that makes it highly measurable.

copies of store ads and pictures of in-store displays. One responsibility of the sales force is to do store checks to verify that stores are doing what they promised. Promotions that contain a response device, such as coupons, have a built-in evaluation measure.

As for consumer promotions, the following evaluation measures are the most popular:

Principle

Proof of performance, various types of responses to promotional tactics, and payout planning provide built-in evaluation measures for promotions.

Measure	% Used
Sales	46
Response rates	20
Awareness	10
Other mix	9
Redemption rates	4

The efficiency of a sales promotion offer can be evaluated in terms of its financial returns more easily than advertising. We compare the costs of a promotion, called a **payout analysis**, to the forecasted sales generated by the promotion. A **break-even analysis** seeks to determine the point at which the total cost of a promotion exceeds the total revenues generated, identifying the point where the promotion is not wise to do. Figure 19.4 depicts this analysis.

Principle

Ultimately, public relations efforts are evaluated in terms of opinion change and relationship tracking.

Public Relations The evaluation of public relations examines the success of getting a message out to a target audience in terms of *output* (materials produced and distributed) and *outcomes* (acceptance and impact of the materials). Outcomes can be measured in terms of changes in public opinion and relationship tracking. The output evaluation is conducted by asking questions such as these: How many placements (news releases that ran in the media) did we get? How many times did our spokesperson appear on talk shows? How much airplay did our public service announcements receive or how much and what kind of buzz are Twitters generating? The results are presented in terms of counts of minutes, mentions, or inches.

Content analysis also helps determine the favorability of the coverage, share of voice, and issue and competitor coverage. Ongoing public opinion tracking studies ask these questions: Has there been a change in audience knowledge, attitudes, or behavior (e.g., pretesting versus post-testing)? Can we associate behavior change (e.g., product trial, repeat purchase, voting, or joining) with the public relations effort? The most common measures of output and outcomes in public relations are summarized here:

Output Objectives Achieved

- **Production** Number of PR messages, such as news releases or brochures, generated.
- **Distribution** Number of media outlets (TV stations, newspapers) receiving PR products.
- **Coverage** Number and size of clips, column inches, or minutes of time or space.
- **Impressions** Media placements multiplied by circulation or broadcast reach.
- **Advertising Value** Equivalent ad costs for time or space.
- **Content Analysis** Positive or valence (whether the story or mention seems to be more positive or negative), key messages (the idea in the story), sources, and prominence.

Outcome Objectives Achieved

- **Awareness** Aided and unaided recall.
- **Attitudes** Perceptions, preferences, and intent to buy.
- **Behavior** Did target do what you wanted them to do?

The search for methods to tie public relations activities to bottom-line business measures, such as ROI (return on investment), is like the quest for the Holy Grail. PR practitioners would like to demonstrate ROI because it would provide even more support for the importance of PR effects. A surrogate ROI measure can be based on shareholder value, which can be seen as a company or brand's reputation capital. For example, research conducted on companies with the most effective employee communication programs has determined that they provide a much higher total return to shareholders. Recently Web-based analytical tools are making it possible to connect earned media results to online business goals, such as generating website traffic, sales leads, revenue, and donations for nonprofit organizations.²³

copies of store ads and pictures of in-store displays. One responsibility of the sales force is to do store checks to verify that stores are doing what they promised. Promotions that contain a response device, such as coupons, have a built-in evaluation measure.

As for consumer promotions, the following evaluation measures are the most popular:

Measure	% Used
Sales	46
Response rates	20
Awareness	10
Other mix	9
Redemption rates	4

The efficiency of a sales promotion offer can be evaluated in terms of its financial returns more easily than advertising. We compare the costs of a promotion, called a **payout analysis**, to the forecasted sales generated by the promotion. A **break-even analysis** seeks to determine the point at which the total cost of a promotion exceeds the total revenues generated, identifying the point where the promotion is not wise to do. Figure 19.4 depicts this analysis.

Public Relations The evaluation of public relations examines the success of getting a message out to a target audience in terms of *output* (materials produced and distributed) and *outcomes* (acceptance and impact of the materials). Outcomes can be measured in terms of changes in public opinion and relationship tracking. The output evaluation is conducted by asking questions such as these: How many placements (news releases that ran in the media) did we get? How many times did our spokesperson appear on talk shows? How much airplay did our public service announcements receive or how much and what kind of buzz are Twitters generating? The results are presented in terms of counts of minutes, mentions, or inches.

Content analysis also helps determine the favorability of the coverage, share of voice, and issue and competitor coverage. Ongoing public opinion tracking studies ask these questions: Has there been a change in audience knowledge, attitudes, or behavior (e.g., pretesting versus post-testing)? Can we associate behavior change (e.g., product trial, repeat purchase, voting, or joining) with the public relations effort? The most common measures of output and outcomes in public relations are summarized here:

Output Objectives Achieved

- **Production** Number of PR messages, such as news releases or brochures, generated.
- **Distribution** Number of media outlets (TV stations, newspapers) receiving PR products.
- **Coverage** Number and size of clips, column inches, or minutes of time or space.
- **Impressions** Media placements multiplied by circulation or broadcast reach.
- **Advertising Value** Equivalent ad costs for time or space.
- **Content Analysis** Positive or valence (whether the story or mention seems to be more positive or negative), key messages (the idea in the story), sources, and prominence.

Outcome Objectives Achieved

- **Awareness** Aided and unaided recall.
- **Attitudes** Perceptions, preferences, and intent to buy.
- **Behavior** Did target do what you wanted them to do?

The search for methods to tie public relations activities to bottom-line business measures, such as ROI (return on investment), is like the quest for the Holy Grail. PR practitioners would like to demonstrate ROI because it would provide even more support for the importance of PR effects. A surrogate ROI measure can be based on shareholder value, which can be seen as a company or brand's reputation capital. For example, research conducted on companies with the most effective employee communication programs has determined that they provide a much higher total return to shareholders. Recently Web-based analytical tools are making it possible to connect earned media results to online business goals, such as generating website traffic, sales leads, revenue, and donations for nonprofit organizations.²¹

Principle

Proof of performance, various types of responses to promotional tactics, and payout planning provide built-in evaluation measures for promotions.

Principle

Ultimately, public relations efforts are evaluated in terms of opinion change and relationship tracking.

Website Evaluation Evaluating the communication effects of website advertising is still a new game. Some performance indicators are traffic volume, such as page views or the simple number of visitors to a site. Banner advertising and other online ads are evaluated using click-through rates, and the one thing the industry has learned is that this form of advertising is decreasing in effectiveness. One problem is that people who have little or no interest in the product may click on the banner by accident or curiosity. Because this type of advertising is sold by pay-per-click (PPC), the advertiser has to pay regardless of the reason for the click. Pop-up banners can get more attention, but they are also seen as more irritating.

Instead of click-through rates, some advertising uses a cost-per-lead metric that records how well the click-through generates prospects, an attempt to get at ROI. The most important metric for Internet advertising, however, is conversion rate, which is the percentage of visitors to a site who complete a desired action, such as playing a game, signing up for a newsletter, or buying something. Of course, online sales are an important measure of a website's effectiveness.

The more sophisticated conversion rate services not only measure how well the site generates action but also how customers navigated the site. This information is obtained by tracking where they come from, what search terms they used, how they move around within the website, and where they go when they leave. It gives a more comprehensive picture of the path visitors and customers use in navigating a site.

Forrester Research has been developing methods for evaluating websites for years. The company tracks effectiveness of website performance in terms not only of making a brand promise, but in terms of also delivering easy-to-find and use information. Harley Manning, a Forrester vice president, reported that his company's research has found that "on average sites did a better job of making the brand promise than they did of meeting customer needs." Manning explains how this research is used to evaluate good and bad performing websites in *The Inside Story*. (Check Forrester's website at www.forrester.com.)

Special Advertising Situations

The various types of advertising identified in Chapter 18 are evaluated using many of the tools discussed previously; however, they each have their own particular objectives that affect how they are evaluated.

The primary objective of *retail advertising* is to generate store traffic. The results of traffic-building promotions and advertising are simple counts of transactions as well as the change in sales volume of brands receiving promotional support. Participation counts can be used to estimate the pull of special events. Loyalty is evidenced through participation in frequent buyer programs and measured in terms of registration, store visits, and average purchases per visit.

In *B2B marketing*, a common form of evaluation is a count of sales leads based on response devices such as calls, e-mails, and cards returned to the advertiser. Another common B2B objective is *conversion*, similar to the conversion rate in the online discussion; this refers to the number of leads that turn into customers who make a purchase. Conversion rates, which are percentages of the leads, are also calculated for most marketing communication tools used in support of B2B programs.

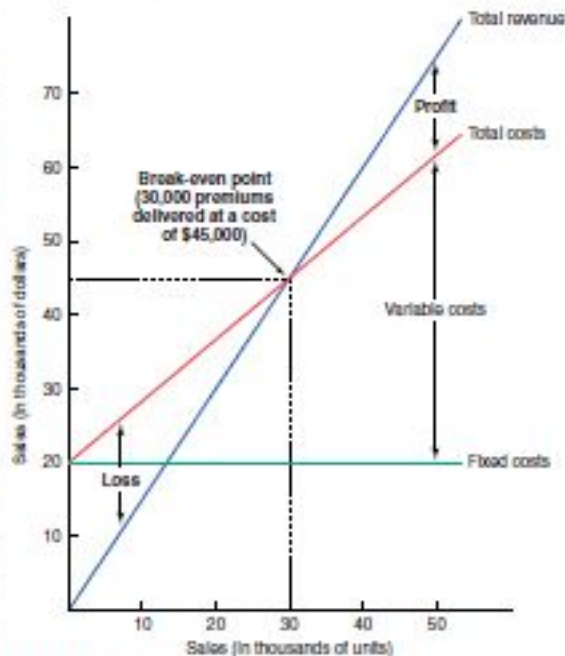


FIGURE 19.4

A Sales Promotion Break-Even Analysis

At the break-even point, where 30,000 premiums are delivered at a cost of \$45,000, the sales revenues exactly cover, but do not exceed, total costs. Below and to the left of the break-even point (in the portion of the diagram marked off by dashed lines), the promotion operates at a loss. Above and to the right of the break-even point, as more premiums are sold and sales revenues climb, the promotion makes a profit.

THE INSIDE STORY

How Web Sites Build Brands (Or Don't)

Harley Manning, Vice President and Research Director, Customer Experience, Forrester Research, Inc.



Brand is at or near the top of the priority list for companies doing business on the Web. In one Forrester survey, decision makers at 148 companies rated building the brand as the second most important goal for their business-to-consumer sites.

The same group said that when considering new content and features for their sites, building the brand was as important as supporting customer goals.

So if online brand building is so important, how can companies make sure that their websites build their brands? To answer that question, we started by interviewing brand strategists at seven of the largest interactive agencies, including Razorfish and Tribal DDB Worldwide. We also talked to marketers at two companies that we respect for their brand management expertise: Johnson & Johnson and Procter & Gamble.

We found broad agreement that websites serve two equally critical functions: They are both a communication medium and a delivery channel. In other words, consumers expect sites to not just make the brand promise but also to keep that promise by providing value in the form of detailed product information that's easy to find and use.

As a result of our initial research, we created a methodology that grades how well sites make the brand promise and keep the brand promise. Our approach is a type of heuristic evaluation, sometimes referred to as an expert review. Standard heuristic evaluations date from the late 1990s and depend on three factors: detailed user descriptions, relevant user goals, and a valid

set of rules (or heuristics) that identify known types of user experience flaws. We stuck with that approach to gauge how well the sites served consumers (i.e., kept the brand promise).

To measure how well sites make the brand promise, we created a variation on the standard expert review methodology. Instead of starting with user descriptions and goals, we begin by identifying the brand attributes of the sites we intend to grade. We do this by looking at their annual reports and the public statements made by their executives. Then we collect above-the-line collateral ranging from annual reports to magazine ads to TV commercials. This preparation lets us check sites for cross-channel consistency of logos, colors, typography, and layout as well as consistency with intended brand attributes such as "reliable" or "innovative" (two of the most common).

We've been using this methodology now for about four years. During that time we've spotted a number of interesting trends. For one, websites on average do a poor job of building either major aspect of brands online: In an analysis of more than 150 reviews we found that only 8% of sites passed both halves of our test.

We also found that, on average, sites did a better job of making the brand promise than they did of meeting customer needs. Sixty sites passed our tests for making the brand promise but only 17 did an adequate job of serving customers' goals. For example, sites from companies as diverse as Nike, Edward Jones, BP, and Coca-Cola excel at online copy and imagery, but fall flat on user experience basics like providing the right content or making that content easy to read via easily legible text.

Like other types of advertising, *nonprofit organizations* need to communicate efficiently and effectively, particularly as they are stewards who must ensure that limited funding is used effectively and efficiently. Evaluation of communication for foundations and nonprofit organizations is an emerging field, and challenges to assessing effectiveness are many. Activities and behaviors being measured are complex and campaigns are often a multi-phase effort. There are no standard and widely accepted guidelines for communication evaluation for the nonprofit sector. Nonetheless, measuring success of a campaign begins with setting objectives and evaluating the progress toward achieving the objectives.²²

International marcom is difficult to evaluate because of market differences (e.g., language, laws, cultural norms) and the acceptability of various research tools. There may also be incompatibilities among various measurement systems and data analysis techniques that make it difficult to compare the data from one market with similar data from another market. An international evaluation program for advertising should focus, at least initially, on pretesting because unfamiliarity with different cultures, languages, and consumer behaviors can result in major miscalculations. Pretesting helps the advertiser correct major problems before miscommunication occurs. To evaluate the global/local question, the best approach is to test two ads that are both based on the global advertising strategy: a locally produced version of the advertising and an original ad produced locally.

Campaign and Program Evaluation

While it is ideal to know how each marketing communication function has performed, the reality in most campaigns is that a variety of functions and media are used to reach and motivate customers to respond. Time is a factor as well. We mentioned in the beginning of this chapter that there is a debate about advertising's ability to impact short-term sales results as well as long-term branding.

University of Southern California Professor Gerard Tellis reminds us that advertising not only has *instantaneous* effects (consumer responds immediately) but also *carryover* effects (delayed impact).²³ Any evaluation of campaign effectiveness needs to also be able to track both types of effects over time. Differential effects also complicate the ongoing evaluation of IMC programs that operate with a 360-degree total communication philosophy, although tracking may be easier in some ways because the evaluation is continuous.

The Synergy Problem The problem with evaluating campaigns—particularly IMC campaigns—is estimating the impact of synergy. Intuitively we know that multichannel communication with messages that reinforce and build on one another will have more impact than single messages from single sources; however, that can be difficult to prove. As Bob Liodice, CEO of the Association of National Advertisers, observes, “there is no single, consistent set of metrics that transcends discipline-centric measurements.”²⁴

If the campaign planning is well integrated, which means each specialty area cooperates with all others in message design, delivery, and timing, then there should be a synergistic effect. This means that the overall results are greater than the sum of the individual functional areas if used separately.

A number of studies have attempted to evaluate IMC impact by comparing campaigns that use two or three tools to see what is gained when more message sources are added to the mix. For example, a study by the Radio Ad Effectiveness Lab reported that recall of advertising is enhanced when a mix of radio and Internet ads are used rather than just website advertising alone.²⁵ A study published in the *Journal of Advertising Research* developed a multidimensional approach based on evaluation of four factors including unified and consistent communication, strategic consistency in targeting (different messages for different audiences), database communication, and relationship programs.²⁶ Such studies of both the platforms and components of IMC are beginning to tease out the effects of synergy, but they are a long way from evaluating the effects of a total communication program.

The most common way of measuring a campaign's total impact is the brand tracking approach mentioned previously. As various ingredients in the campaign are added and taken away, changes in tracking study results can show the effects and help identify what combinations of marketing communication functions and media work best for a brand. In other words, has the brand become stronger on critical dimensions of the image, such as personality and positioning cues, because of the campaign?

A complication in evaluating programs is the need to consider other messages and contact points beyond traditional marketing communication. As we have said, in a 360-degree total communication program, brand experiences, such as those involved with customer service and WoM, may be even more important than traditional marketing communication. A *Principled Practice* feature by Professor Keith Murray demonstrates how these unconventional message effects can be measured.

Connecting the Dots The challenge in campaign and program evaluation is to pull everything together and look at the big picture of campaign performance rather than the individual pieces and parts. One of the first places to start is defining the objectives—all the various effects—and then adequately and realistically measure the campaign's performance against those objectives. Here's an example that demonstrates how evaluation methods can be matched to the original campaign objectives.

Effie award winner UPS wanted to reposition itself by broadening its package delivery image.²⁷ Although UPS owned ground delivery, it lost out to Federal Express in the overnight and international package market. UPS knew from its customer research that to break out of the “brown and ground” perception, the company had to overcome the inertia of shipping managers who use UPS for ground packages and FedEx for overnight and international. The company also had to shift the perception of senior executives from a company that handles packages to a strategic partner in systems planning. From these insights came three sets of objectives that focused on

A PRINCIPLED PRACTICE

Can a Broken Guitar Really Hurt United?

Keith Murray, Associate Dean, College of Business, Bryant University



It's been a YouTube hit and made the e-mail rounds for a long time—the "United Breaks Guitars" video. It's the story of a musician who made a trip on United Airlines (UA) and was compelled to "check" his Taylor guitar as baggage—which is how

UA came to be vulnerable to the charge that, in the handling of it, they damaged the instrument.

When asked to make it right—to pay for fixing the Taylor guitar—UA declined.

The musician, Dave Carroll, responded by writing a ballad about the experience and posting it on YouTube and the Internet. (See it at www.youtube.com/watch?v=5YGo4zOqozo.)

A lot of people viewed the clip; it's catchy and tells the story in an amusing way. People have sent it to their friends . . . and friends have sent it to their friends, which has led now to literally millions of people seeing it! You can also see a CBS report about the incident at www.youtube.com/watch?v=PGNtQF3n8/Y&feature=related.

So was UA smart to ignore this incident? In other words, did UA really "save" the \$1,200 it would have taken to fix the guitar in the first place?

It seems clear that UA might have been short sighted. If you "do the numbers," you come to the conclusion that UA may have paid a much higher price than it realized!

The table following shows how much money might have been—and is still being?—lost by UA from 8 million people (in 2010) seeing the YouTube clip and deciding against using UA. It gives various levels of impact from 1 to 10 percent. It also com-

pares the percent of this lost revenue against UA's sales revenue in 2008.

# of \$500 Trips NOT Taken In a Year	Percent of Viewers Influenced by YouTube Videos to NOT Patronize UA in a 12-Month Period			
	1%	2%	5%	10%
One trip	\$30M	\$60M	\$150M	\$300M
% '08 revenue	0.15	0.31	0.78	1.56
Two trips	\$60M	\$120M	\$300M	\$600M
% '08 revenue	0.31	0.63	1.56	3.13

What the calculations show is compelling. If only 1 percent of those who learned of the broken Taylor guitar were affected by the story, then UA only lost somewhere between \$30 million and \$60 million. However, if the negative influence is higher, then UA could arguably have forgone as much as half a billion dollars, or about 3 percent of its annual sales! In any case, all of these figures stand in stark contrast to the \$1,200 asked for by Mr. Carroll in the first place!

This gets to the real point of the story: if UA (or any company) has a flawed system to handle and remedy customer complaints—in other words, if UA has customer service "issues" that produce unsatisfactory results for customers—then it can pay a very high price for its poor service. And the damage is exponential, because the average person tells about 10 people about a bad brand experience—all of which has a chilling effect on patronage by those who hear such tales of woe!

These numbers show the huge impact of failure to pay attention to customer complaints and service system problems. Can one little broken Taylor guitar—and all the other little failures each day—affect a mammoth company like United Airlines? You bet.

breaking through awareness, breaking the inertia trance, and breaking the relevance trance. Here's how the campaign performed on those objectives. Notice the mix of perception, image, and behavioral measures.

Objective 1: Breaking through Awareness

- Awareness of the Brown campaign outpaced all past UPS advertising measured in the 10-plus-year history of its brand tracking study.
- Among those aware of the campaign, correct brand linkage to UPS was 95 to 98 percent across all audiences (compared to a historical average of 20 to 40 percent for past UPS advertising).

- “What Can BROWN Do for You?” has taken hold in popular culture. For instance, the tagline was mentioned in both *Saturday Night Live* and *Trading Spaces* shows.

Objective 2: Breaking the Inertia Trance

- With shipping decision makers, the brand showed steady and significant gains in key measures like “Helps my operation run more smoothly,” “Dynamic and energetic,” and “Offers a broad range of services.”
- International shipping profitability increased 150 percent, and overnight volume spiked by 9.1 percent after the campaign ran. The targeted companies’ total package volume increased by 4.39 percent.
- From the start of the campaign in March to the year-end, annual ground shipping revenue grew by \$300 million.
- The campaign was a hit in terms of response with a 10.5 percent response rate and an ROI of 1:3.5. In other words, every dollar spent on the campaign generated \$3.5 dollars in revenue.

Objective 3: Breaking the Relevance Trance

- For the first time in the 10-plus-year history of the brand tracking study, UPS leads FedEx in all image measures among senior-level decision makers. All significant brand image measures continued upward.
- Among senior decision makers, the biggest gains were in key measures like “For people like me,” “Acts as a strategic partner to my company,” “Helps in distribution and supply chain operations,” and “Provides global competitive advantage.”
- At the start of the campaign, annual nonpackage (supply chain) revenue was approximately \$1.4 billion. By the end of the year, nonpackage revenue had almost doubled to \$2.7 billion.

Bringing It All Together Beyond connecting the objectives to the measurements, advertisers continue to search for methods that will bring all the metrics together and efficiently and effectively evaluate—and predict—brand communication effectiveness. A small Florida agency, Zimmerman Advertising, positions itself specifically on that issue. It promises to deliver long-term brand building as well as short-term sales. The president explains, “The biggest problem marketers face today is connecting advertising to retail sales.” To help determine the impact of its ads on sales, Zimmerman measures ad response immediately using a toll-free number and proprietary software that tracks sales to specific ads.²⁸

The giant media research company, Nielsen, has undertaken a project to find the links in its various marketing metrics. In particular, the company wants to connect its Nielsen Media Research TV ratings with its A.C. Nielsen retail scanner and consumer panel data. Then it wants to mix in data from its Nielsen/NetRatings for online advertising and Nielsen Buzzmetrics online-buzz measurement system.²⁹ This is just one example of how important it is to connect the dots to draw an accurate picture of consumer response to marketing communication.

Ultimately, the goal is to arrive at holistic, cross-functional metrics that are relevant for integrated communication, a task undertaken by Dell Computers and its agency DDB. Given Dell’s direct-marketing business model, the company had extensive call and order data in its database. DDB helped organize the collection of detailed marcom information, which made it possible to begin linking orders to specific marcom activities. This new marcom ROI tracking system made it possible for Dell to recognize a 3 percent gain in the efficiency of its marketing communication. As the metrics system became more sophisticated, it also began to move from a reporting and metrics evaluation engine to a strategic tool providing deeper insights into consumer behavior.³⁰

Many pieces are still missing in the evaluation of advertising, not to mention more complex IMC programs. Research think tanks are struggling to find better ways to measure consumers’ emotional connections to brands and brands’ relationships with their customers³¹ and how those connections and relationships are affected by various types of marketing communication messages. But let’s end with an inspirational video called “Life Lessons from an Ad Man.” It’s a little old, but Rory Sutherland makes a good point about having fun even as you contemplate how marketing communication creates intangible value for brands. See it at: www.ted.com/talks/rory_sutherland_life_lessons_from_an_ad_man.html.