18 Measuring the Effectiveness of the Promotional Program

Source: Truth Initiative

Learning Objectives

LO 18-1 | Compare reasons for and against measuring the effectiveness of promotional programs.

LO 18-2 | Describe the tools and processes available for assessing promotional program effectiveness.

LO 18-3 | Discuss the limitations of current methods for measuring advertising effects.

LO 18-4 | Compare different methods of measuring effectiveness of other promotional programs.

The David Ogilvy Awards for Advertising Effectiveness

One of the most prestigious awards an advertising agency and its client can receive is an Ogilvy Award. The award is named for longtime advertising executive David Ogilvy, who was always a strong advocate of the use of research in advertising. Sponsored by the Advertising Research Foundation (ARF), the awards are given for excellence in advertising research and/or the creative use of research in the advertising development process. While initially the awards were purely advertising related, awards are now given in a variety of communications areas.

ARF Grand Ogilvy Award Winner

Mr. Clean: "Cleaner of Your Dreams"

In case you don't know who Mr. Clean is, ask one of your parents. For years, Mr. Clean was one of the leading brands of cleaning products, characterized by the fictional character Mr. Clean. However, over time, the brand hit the decline stage of the product life cyle, and the company (Procter & Gamble) and its agency knew something had to be done. Research uncovered a surprising and interesting connection between sex and cleaning that ended up returning the brand into the minds, hearts, and homes of today's modern cleaners. The research revealed the insight that "there's nothing sexier than a man who cleans," which led to the idea that "Mr. Clean is the cleaner of your dreams." The campaign struck the right balance between using the Mr. Clean icon as entertainment and establishing the brand's superior cleaning credentials.

Why did this campaign win?

- The creative team translated the insights into a breakthrough creative ad.
- Campaign research followed best practice methodology.

- The research identified the need to connect the past and make it more modern.

Category: Breaking Through
Gold: Farmers Insurance: We Know from Experience

The insurance category is one of the biggest advertising spenders in America, with companies spending more than $5 billion a year on ads. Farmers Insurance was a bit player, spending less than one-tenth what giants Progressive and Geico spent. In order to compete, Farmers realized it needed a strategy that was attention getting and engaging. To do so, it set out on a comprehensive research program with a differentiating and breakthrough brand idea.

By showing a variety of real insurance claims that Farmers agents had actually handled, Farmers expressed the value of experience and knowledge as invaluable assets for an insurance company. The tagline "We've seen a thing or two, so we know a thing or two" was a big hit, and the campaign increased awareness, differentiated Farmers from other insurance companies, and increased consideration.

It is estimated that the average American consumer sees approximately 30 insurance ads a week. The category includes the highest spending advertiser (Geico). Farmers Insurance, a respected company with almost 90 years of experience, ranks ninth in category spending. In addition, the brand charges a premium, compared to its competitors, in a price-competitive environment. Using Bayesian modeling and waves of qualitative research, it was determined that the attributes of smart, trustworthy, knowledgeable, and experienced were most important to those willing to pay more for insurance than they would have to by using the price competitors.

The campaign sought to demonstrate Farmers' capability to handle every problem, regardless of how outrageous or unlikely, featuring more than 30 different unbelievable claims, each based on a real insurance claim that a Farmers agent handled. These crazy accidents were brought to life in video channels through Professor Burke (J.K. Simmons), the tweed-clad teacher/spokesman, who reenacted the incidents to customers via exhibits in a museum-like environment, the Hall of Claims. These insurance oddities also came alive on Farmers.com—the key portal for people considering new insurance and quotes—and engaging digital campaigns, also based on real claims. During Oscar season, a new award show called "The Burkies" was introduced, and it rolled out six trailers for six fake movies, all based on real Farmers insurance claims.

page 594

Category: Unexpected Targeting
Gold: truth: Stop Profiling

When truth®, a national, mass media, youth smoking-prevention campaign, launched in 2000, 23 percent of all U.S. youth smoked cigarettes, and smoking continued to be a top public health concern. truth soon established itself as a highly successful anti-tobacco campaign by exposing the deceptive practices of the tobacco industry to teens open to smoking and encouraging them to "rebel against the industry" and reject tobacco. A "truth Teen" was an edgy thought leader, a dreamer, a rebel, and yet was judicious and conscientious enough to make the right health decisions. By 2014, the challenge had evolved. At that time only 9 percent of youths smoked cigarettes and, by 2016, teen smoking reached a historic low of 6 percent. While a highly successful campaign, with battles being won, the war was not over. For example, low-income individuals were still more than twice as likely to smoke.

Research has shown that this generation of teens and young adults are passionate and enthusiastic about expressing themselves through social movements. Given the declining tobacco use among young people and this generation's interest in social movements, the truth campaign was repositioned as a social movement to end smoking—to FinishIt. By again focusing on Big Tobacco's profiling tactics, the campaign successfully reframed the issue of smoking as a social injustice that would be compelling not just to the specific population subgroups targeted, but to all youth and young adult nonsmokers and smokers as well.

A multimethod research approach was implemented to present the content of the public education campaign and to evaluate its impact among the target audience. Premarket concept and advertisement (ad) testing was employed to assess receptivity, message comprehension, and emotional response, with ad tracking used to assess recall over time.

The campaign ignited young adults' sense of justice, sparking outrage at Big Tobacco while successfully driving down youth smoking. As noted by agency Truth Alliance:

- The campaign has generated more than 98.5 million video views across all of truth's social media channels and nearly 2 milion social engagements, including social support from stakeholders such as the American Public Health Association and NYC Pride. It also sparked conversations around a long-term partnership with the Human Rights Campaign.
- truth experienced a nearly 700 percent increase in website traffic in the seven days following this campaign.
- In 2017, truth enlisted more than 200,000 + new "Finishers" in the movement to end smoking for good at thetruth.com.
- Media outreach for the campaign generated nearly 120 placements (302.2 million earned media impressions to date) including top-tier placements in *The Washington Post, Mashable, Fast Company, HowStuffWorks, Essence, Bustle, Huffington Post, Teen Vogue, CheddarTV, Hello Giggles,* and *ATTN.*

Perhaps most important, according to Truth Alliance, was that "it actually changed beliefs about young people's attitudes towards Big Tobacco, susceptibility to smoke, intention to smoke and smoking behavior."

Category: Changing Consumer Behavior
Gold: Rice Krispies Treats: So Much to Love

As American children prepared to head back to school in the fall, Kellogg's Rice Krispies Treats found itself with a declining market share compared to other lunchtime snacks, and brand awareness and consideration figures continued to fall. To reverse this trend, Kellogg's and its partner agency, Leo Burnett, set off to find a new way to make a strong connection with parents. Using Google Tools, the research team uncovered an opportunity in the search behaviors of moms and dads preparing their children for another year of school. Analyzing trillions of Google.com search queries, the findings showed that a clear linkage existed between “school snacks” and “lunchbox notes.” Kellogg’s capitalized on the opportunity by designing packaging for Rice Krispies Treats that provided parents with a heart-shaped empty space to write notes to their children. To get the message out, the campaign called “So Much to Love” captured the anxiety, excitement, and love associated with the back-to-school season. Consumers clearly found the spots compelling. During the campaign, Google and YouTube registered huge lifts in search volume for the product keywords as video ads drove consumers to search to learn more about Rice Krispies Treats write-on packaging.

These few examples are very useful in demonstrating the value of research to advertisers and their agencies. There are many more such examples of the use of research in a variety of causes and marketing strategies, including consumer behavior, segmentation and targeting, and education and community change. The awards are sponsored by the Advertising Research Foundation (ARF), and these interesting case studies are provided in more detail at the website www.thearf.org. It is worth a visit to read these case write-ups of the best in research for advertising purposes.

Source: The David Ogilvy Awards, www.thearf.org.

As noted throughout this text, the increased emphasis on accountability is forcing many companies to evaluate, or reevaluate, their IMC plans. Both clients and agencies are continually striving to determine whether their communications are working and how well they are working relative to other options. A number of studies indicate that marketing managers are not confident in the metrics they are currently using and are searching for new methods designed to assist in this endeavor. Companies and organizations continue to work together in an attempt to provide answers to these questions and to develop new ways to measure communications effectiveness.

Measuring the effectiveness of the promotional program is a critical element in the IMC planning process. Research allows the marketing manager to evaluate the performance of specific program elements and provides input into the next period’s situation analysis. It is a necessary ingredient to a continuing planning process, yet it is often not carried out.

In this chapter, we discuss some reasons firms should measure the effectiveness of their IMC programs, as well as why many decide not to. We also examine how, when, and where such measurements can be conducted.

Most of our attention is devoted to measuring the effects of advertising because much more time and effort have been expended developing evaluation measures in advertising than in the other promotional areas. We will, however, discuss measurement in other areas of the IMC program as well. You'll recall that we addressed the methods used to evaluate many of the other promotional elements in previous chapters.

It is important to understand that in this chapter we are concerned with research that is conducted in an evaluative role—that is, to measure the effectiveness of advertising and promotion and/or to assess various strategies before implementing them. This is not to be confused with research discussed earlier in the text to help develop the promotional program, although the two can (and should) be used together, as you will see. While evaluative research may occur at various times throughout the promotional process (including the development stage), it is conducted specifically to assess the effects of various strategies. We begin our discussion with the reasons effectiveness should be measured as well as some of the reasons firms do not do so.

ARGUMENTS FOR AND AGAINST MEASURING EFFECTIVENESS

Almost anytime one engages in a project or activity, whether for work or for fun, some measure of performance occurs. In sports, you may compare your golf score against par or your time on a ski course to other skiers' performances. In business, employees are generally given objectives to accomplish, and their job evaluations are based on their ability to achieve these objectives. Advertising and promotion should not be an exception. It is important to determine how well the communications program is working and to measure this performance against some standards.

Reasons to Measure Effectiveness

LO 18-1

Assessing the effectiveness of ads both before they are implemented and after the final versions have been completed and fielded offers a number of advantages:

1. *Avoiding costly mistakes.* The top three advertisers in the United States were expected to spend over $11 billion in advertising and promotion in 2018. The top 10 spent a total of over $25 billion.[1] This is a lot of money to be throwing around without some understanding of how well it is being spent. If the program is not achieving its objectives, the marketing manager needs to know so he or she can stop spending (wasting) money on it.

   Just as important as the out-of-pocket costs is the opportunity loss due to poor communications. If the advertising and promotional program is not accomplishing its objectives, not only is the money spent lost but so too is the potential gain that could result from an effective program. Thus, measuring the effects of advertising does not just save money. It also helps the firm maximize its investment. Think about how many times you have seen an ad and had no idea what it was trying to communicate!

2. *Evaluating alternative strategies.* Typically a firm has a number of strategies under consideration. For example, there may be some question as to the degree to which each medium should be used or whether one message is more effective than another. Or the decision may be between two promotional program elements. A key issue for the manager is to be able to determine how effective each one was. As you have seen in previous chapters, numerous companies have also reallocated monies to nontraditional media, and studies have revealed that advertisers said they planned to spend more monies in digital and social media in the coming years. The question is, should research be spent on traditional, social, or digital advertising or other media and in what combination? Research may be designed to help the manager determine which strategy is most likely to be effective.

3. *Increasing the efficiency of advertising in general.* You may have heard the expression "can't see the forest for the trees." Sometimes advertisers get so close to the project they lose sight of what they are seeking, and because they know what they are trying to say, they expect their audience will also understand. They may use technical jargon that not everyone is familiar with. Or the creative department may get too creative or too sophisticated and lose the meaning that needs to be communicated. How many times have you seen an ad and asked yourself what it was trying to say, or how often have you seen an ad that you really like, but you can't remember the brand name? Conducting research helps companies develop more efficient and effective communications. An increasing number of clients are demanding accountability for their promotional programs and putting more pressure on the agencies to produce. As the lead-in to this chapter discussing the Ogilvy Award winners demonstrates, effective research can be used for both of these purposes. The way to conduct this research is not so clear, as can be seen in Digital and Social Media Perspective 18–1).
4. *Determining if objectives are achieved.* In a well-designed IMC plan, specific communication objectives are established. If objectives are attained, new ones need to be established in the next planning period. An assessment of how program elements led to the attainment of the goals should take place, and/or reasons for less-than-desired achievements must be determined. Research should address whether the strategy delivers the stated objectives and how appropriate the measures used to make this assessment are.

Reasons Not to Measure Effectiveness

While it seems obvious that it makes sense to measure effectiveness, the fact remains that in too many instances this is not done. Whereas advertisers know that it is important to measure effectiveness, with as many as 90 percent considering it a priority, many do not do so, or if they do, they are not confident of the results.

Companies give a number of reasons for not measuring the effectiveness of advertising and promotions strategies:

1. *Cost.* Perhaps the most commonly cited reason for not testing (particularly among smaller firms) is the expense. In one of the surveys cited, it was noted that while some companies spend as much as 25 percent of their revenue on marketing and advertising, 70 percent of them spend less than 2 percent on measuring effectiveness.[2] Good research can be expensive in terms of both time and money. Many managers decide that time is critical and they must implement the program while the opportunity is available. Many believe the monies spent on research could be better spent on improved production of the ad, additional media buys, and the like.

   While the first argument may have some merit, the second does not. Imagine what would happen if a poor campaign were developed or the incentive program did not motivate the target audience. Not only would you be spending money without the desired effects, but the effort could do more harm than good. Spending more money to buy media does not remedy a poor message or substitute for an improper page 597 promotional mix. For example, one of the nation's leading brewers watched its test-market sales for a new brand of beer fall short of expectations. The problem, it thought, was an insufficient media buy. The solution, it decided, was to buy all the TV time available that matched its target audience. After two months, sales had not improved and the product was abandoned in the test market. Analysis showed the problem was not in the media but rather in the message, which communicated no reason to buy. Research would have identified the problem, and millions of dollars and a brand might have been saved. The moral: Spending research monies to gain increased exposure to the wrong message is not a sound management decision.
2. *Research problems.* A second reason cited for not measuring effectiveness is that it is difficult to isolate the effects of promotional elements. Each variable in the marketing mix affects the success of a product or service. Because it is often difficult to measure the contribution of each marketing element directly, some managers become frustrated and decide not to test at all. They say, "If I can't determine the specific effects, why spend the money?" In a study conducted by the Fournaise Marketing Group—a marketing performance and measurement

business—reviewing over 500 marketing campaigns, briefs, and effectiveness reports, it concluded that the metrics being used were "shocking" in a very negative sense.[3] Simply put, the report concluded that marketers did not understand how to measure effectiveness.

This argument that not being able to determine specific effects, so no measurement should take place, suffers from weak logic. While we agree that it is not always possible to determine the dollar amount of sales contributed by promotions, research can provide useful results and, as shown throughout this text, most have useful and specific metrics to evaluate their performance.

3. *Disagreement on what to test.* The objectives sought in the promotional program may differ by industry, by stage of the product life cycle, or even for different people within the firm. There are numerous ways to measure these and not always a consensus as to what measure should be used. The sales manager may want to see the impact of promotions on sales, top management may wish to know the impact on corporate image, and those involved in the creative process may wish to assess recall and/or recognition of the ad. The metrics used to measure effectiveness in traditional media are often quite different than those for digital media. Lack of agreement on what to test often results in no testing at all. A study conducted by the 4As and ANA revealed that many marketers were dissatisfied with their efforts to integrate traditional and digital media, and there is a need to develop appropriate metrics for doing so. At the same time there is much disagreement as to what to measure. As shown in Chapter 15, a variety of metrics are used in measuring the effectiveness of social media alone, and these measures continue to change.

   Again, there is little rationale for this position. With the proper design, many or even all of the above might be measured. Since every promotional element is designed to accomplish specific objectives while contributing to the overall program, research can be used to measure its effectiveness in doing so.

4. *The objections of creative.* It has been argued by many (and denied by others) that the creative department does not want its work to be tested and many agencies are reluctant to submit their work for testing. This is sometimes true. Ad agencies' creative departments argue that tests are

not true measures of the creativity and effectiveness of ads; applying measures stifles their creativity; and the more creative the ad, the more likely it is to be successful. They want permission to be creative without the limiting guidelines marketing may impose. The Chiat/Day ad shown in Exhibit 18–1 reflects how many people in the advertising business feel about this subject. The 1984 commercial was designed by agency Chiat Day to introduced the Apple MacIntosh. It was shown only one time in the United States—during the 1984 Super Bowl. Due to its creativity, the ad is considered by many to be the greatest commercial ever made and a large reason the Mac was so successful.

page 598

EXHIBIT 18–1

In this ad CHIAT/DAY expresses its opinion of recall tests by offering a case history of the 1984 Apple Commercial.

Source: TBWA\CHIAT\DAY

At the same time, the marketing manager is ultimately responsible for the success of the product or brand. Given the substantial sums being allocated to advertising and promotion, it is the manager's right, and responsibility, to know how well a specific program—or a specific ad—will perform in the market. As you have seen throughout this text, managers are placing much more emphasis on social and digital media at the expense of traditional strategies—not just media but creative strategies as well, much to the chagrin of creative directors. A very interesting study by comScore ARS measured the contribution that effective creative versus other marketing variables—ad quality, the media plan, and other elements (price, promotions, distribution, etc.)—made to campaign effectiveness (Figure 18–1).[4]

FIGURE 18–1

Effective Creative Does Contribute to Campaign Effectiveness

Variables	Contribution to Sales Changes (Percent)*
Ad quality†	52%
Media plan‡	13
Other (price, promotion, distribution, etc.)	35

* Numbers represent the percent variance in market share shifts explained by the corresponding factors.

†Ad quality represents the quality of creative based on the ARS Persuasion Score, which measures changes in consumer preference through a simulated purchase.

‡Media plan includes variables such as GRPs, wearout, and continuity/flighting of airing.

Source: comScore ARS Global Validation Summary.

5. *Time.* A final reason given for not testing is a lack of time. Managers believe they already have too much to do and just can't get around to testing, and they don't want to wait to get the message out because they might miss the window of opportunity. Planning might be the solution to the first problem. An article in *Advertising Age* indicates that time pressures are actually on the increase, particularly when it comes to digital. The speed to get content out in the digital age puts a spotlight on previous processes used to validate ad concepts before they are put into the market. As a result, many marketers are eschewing traditional copy testing methods and taking the risk that their ads will succeed on a hunch or gut feel. Although many managers are overworked and time poor, research is just too important to skip.

Digital and Social Media Perspective 18–1 > > >

The Lack of Trust in Effectiveness Measures Inhibits Their Use

It has always been a struggle for managers to get an idea of how well their advertising programs were working. To make matters worse, when they have taken efforts to determine the effectiveness of their programs, they often got pushback from their ad agencies, which argued (and still do!) that it is difficult to measure how ads are contributing to the overall marketing program, and that measuring could actually lead to *lower* effectiveness by stifling creativity. But then, shareholders and others arguing for some of the monies being spent on advertising began to have an impact and marketers started to pay more attention to metrics that might provide them with data to make better management decisions. With the rapid adoption of new digital media, it seems that marketers are (or at least some feel that they are) regressing to a time long ago because it is so difficult to measure the contribution that these new media are making to the overall marketing program. Add to this the fact that there are often more time pressures when placing ads in these media, and the problem increases.

Over the years advertisers came to an agreement as to what measures should be used to measure the effectiveness in traditional media. However, there is no such consensus with social and digital at this point, and there are still many companies that use new media metrics only to evaluate their IMC programs, which is insufficient. At the same time, there are numerous studies that show that managers say they do not feel that they have a good understanding of marketing effectiveness in either channel. Agencies are even more skeptical, believing that their clients' understanding of the effects range from nonexistent to "poor."

- *Click-through rate (CTR)*—Perhaps the most used metric (though there are signs this number is starting to decrease) to measure the effectiveness of display ads on the Internet is the click-through rate. For years, CTR has remained at or about 0.10 percent, but expenditures in digital continue to climb. In 2017, U.K. company dunnhumby.com conducted a study of more than 60 digital advertising campaigns and found no relationship between CTR and sales. In a second study 18 months later, the results were the same. As noted by Jake Ferguson, media planner at dunnhumby, focusing on clicks will have no tangible impact on one's digital media performance and could even be promoting bad campaigns due to a flawed measure.

- *Video ad views*—Video ad revenues make up more than half of Twitter's ad revenues, half of Snapchat's, and 30 percent of Facebook's. Yet there is no consistency among them as to what constitutes a view. For Facebook and Instagram, it is three seconds. For YouTube, it is around 30 seconds. Other social media sites have adopted the Media Rating Council (MRC) standard of "when 50% of the ad's pixels are visible on the screen for two consecutive seconds." So if you are buying an ad on these platforms, you are dealing with different standards for each. In other words, it is not apples to apples.
- *Television*—Artie Bulgrin, executive vice president of insights and strategy at research firm Media Science, says television needs to find a fresh set of measurement metrics if it hopes to maintain its value to advertisers. As television faces increased competition from digital, he says, gaps in knowledge about how to measure effectiveness are undermining advertisers' confidence in the medium.

Measurement companies are being developed to compete against Google Analytics, as a result of recent privacy issues due to the huge amount of data the company has about consumers. And the disagreement goes on. For example, one article appears contending that online metrics have replaced traditional measures of knowing the effect of ads on consumers in the purchase funnel. Another study also demonstrates that online metrics can be useful for this purpose. But then there are others, like Shane Snow, who argue that marketers are getting it all wrong, and that "the next big trend in media is going to require us to rethink publishing metrics entirely." Jack Marshall contends that attempts to measure native advertising are "all over the map." Some managers have even argued that since different companies have different objectives for their campaign, metrics are superfluous. In a survey conducted by the Association of National Advertisers, many managers said they were still trying to figure out exactly what they should measure. In reviewing the survey, the ANA's executive vice president, Bill Duggan, noted that the lack of standardized metrics may hamper future growth.

metamorworks/Shutterstock

To make things even more complicated, there is a growing interest in measures that don't neatly fit into either the digital or traditional category. An article in *Advertising Age* says the future of measurement may not come from surveys but from smiles and/or smirks. The article notes that if you want to know how people react to an ad you need to measure how they feel, not ask them. A number of new companies now offer sophisticated "facial recognition" technologies. For example, Kellogg Company, Mars, Inc., and Unilever have used the services of Affectiva to measure and analyze moment-to-moment facial expressions of people watching videos on a smartphone or laptop. Affectiva's CEO Nicholas Langeveld says that "people have a really hard time articulating their feelings. And sometimes there is a subtle fleeting little emotion that people aren't even aware is happening." Facial recognition will capture these emotions, supporters say. Research company Kantar Millward Brown now offers facial recognition as part of its ad testing suites, and Nielsen and others do as well. A competitor of Affectiva is Emotient, which has roots at the Machine Perception Lab at the University of California, San Diego. Emotient conducts facial recognition studies in controlled groups but is increasing its emphasis on testing ads "in the wild"—that is, analyzing the facial expressions of people viewing ads on digital signs in real life.

So what's the answer? As you will see in this chapter, marketers are trying a variety of methods to try to find out. Many of these measures are new, some have been around for a while, and some are a combination of old and new. New techniques like eye tracking, facial recognition, and neuroscience applications are being used alongside traditional recall and attitude measurement techniques. More and more retailers are using attribution modeling to attempt to determine how various media touch points are performing, and there has been an increase in measuring the effectiveness of emotional versus cognitive ads, the impact of creativity, storytelling versus hard sell, and so on.

Whereas marketers have disagreed on both the value of measuring effectiveness and how to measure for a long time, few these days argue against the need to determine the former. There also seems to be a movement to employ various methods at the same time. As noted by both Affectiva and Kantar Millward Brown, facial recognition should not be used in isolation, but as a complement to traditional copy testing methods. Maybe we can all agree to that?

Sources: Katharine Schwab, "It's Time to Ditch Google Analytics," February 1, 2019, www.fastcompany.com; "TV Needs a Measurement Update," March 29, 2019, www.warc.com; AmyGesenhues, "What Counts as a Video View? A Refresher on How Social Platforms Calculate Video Ad Views," March 29, 2019, www.thirddoormedia.com; Jake Ferguson, "Click-Through Rate Is Well Past Its Sell-by-Date," February 26, 2019, www.warc.com; "Facial-Recognition Lets Marketers Gauge Consumers' Real Responses to Ads," May 18, 2015, www.adage.com; Shane Snow, "Brands Are Measuring Their Content All Wrong," April 22, 2014, www.contently.com; Jack Marshall, "Native Advertising Measurement Is All over the Map," April 17, 2015, www.wsj.com; "Display Ad Effectiveness in the Purchase Funnel," October 22, 2015, www.msi.org; www.mediapost.com; Laurie Sullivan, "Retailers Use Attribution Modeling to Measure the Touchpoints Driving Sales," May 28, 2013, www.emarketer.com.

The second argument can also be overcome with proper planning. Responding to these needs, at least one research agency is in the process of developing new tools that will allow marketers to test different versions of creative ideas and get results back within hours. While timeliness is critical, getting the wrong message out is of little or no value and may even be harmful. There will be occasions where market opportunities require choosing between testing and immediate implementation. But even then some testing may help avoid mistakes or improve effectiveness.

CONDUCTING RESEARCH TO MEASURE ADVERTISING EFFECTIVENESS

What to Test

LO 18-2

We now examine how to measure the effects of communications. This section considers what elements to evaluate, as well as where and how such evaluations should occur.

In Chapter 5 we discussed the components of the communication model (source, message, media, receiver) and the importance of each in the promotional program. Marketers need to determine how each is affecting the communication process. Other decisions made in the promotional planning process must also be evaluated.

Source Factors An important question is whether the spokesperson being used is effective and how the target market will respond to him or her. Or a product spokesperson may be an excellent source initially but, owing to a variety of reasons, may lose impact over time. As shown in Chapter 17, negative publicity can easily change the value of a source. The fact that so many of the companies using Lance Armstrong as a spokesperson terminated their contracts with him was based on the expectation that the target audiences would no longer have positive perceptions of him. The list of celebrities who have fallen out of favor is a long one. Toy manufacturer Hasbro has developed their own methodology for testing the effectiveness of its influencers.

Message Variables Both the message and the means by which it is communicated are bases for evaluation. For example, in the beer example discussed earlier, the message never provided a reason for consumers to try the new product. In other instances, the message may not be strong enough to pull readers into the ad by attracting their attention or clear enough to help them evaluate the product. Sometimes the message is memorable but doesn't achieve the other goals set by management. For example, one study examined what effect sexually themed print ads would have on viewers. Among the numerous results was that men favor sex appeals more than women do and that recall of the brands was lower for sexual ads than for nonsexual ones. Whereas men responded that sexual ads have "high stopping power" for them, their lower brand recall seems to indicate that they are paying more attention to other aspects of the ad than the marketers would prefer.[5] A number of companies are now attempting to determine the specific reactions that viewers have to product and brand messages, including excitement, engagement, stress, and anxiety responses.

Media Strategies Research may be designed in an attempt to determine which media class (for example, broadcast versus print), subclass (newspapers versus magazines), or specific vehicles (which newspapers or magazines) generate the most effective results. Likewise, how does one digital medium compare relative to others, or to traditional media? Perhaps most importantly, how does each medium contribute to the achievement of overall IMC objectives?

An important factor is the **vehicle option source effect,** "the differential impact that the advertising exposure will have on the same audience member if the exposure occurs in one media option rather than another." People perceive ads differently depending on their context.[6]

Another factor to consider in media decisions involves scheduling. The evaluation of flighting versus pulsing or continuous schedules is important, particularly given the increasing costs of media time. As discussed in Chapter 10, there is evidence to support the fact that continuity may lead to a more effective media schedule than does flighting. Likewise, there may be opportunities associated with increasing advertising weights in periods of downward sales cycles or recessions. The manager experimenting with these alternative schedules and/or budget outlays should attempt to measure their differential impact.

As more and more companies and organizations move toward an integrated media mix, it becomes increasingly important to attempt to determine the individual contributions of various media as well as their synergistic effect. As you will see later in this chapter, progress is being made in this regard, but making such a determination is not a simple task.

Budgeting Decisions A number of studies have examined the effects of budget size on advertising effectiveness and the effects of various ad expenditures on sales. Many companies have also attempted to determine whether increasing their ad budget directly increases sales. This relationship is often hard to determine, perhaps because using sales as an indicator of effectiveness ignores the impact of other marketing-mix elements. More definitive conclusions may be possible if other dependent variables, such as the communications objectives stated earlier, are used.

When to Test

Virtually all test measures can be classified according to when they are conducted. **Pretests** are measures taken before the campaign is implemented; **posttests** occur after the ad or commercial has been in the field. A variety of pretests and posttests are available to the marketer, each with its own methodology designed to measure some aspect of the advertising program. Figure 18–2 classifies these testing methods.

IGURE 18–2

Classification of Testing Methods

Pretests		
Laboratory Methods		
Consumer juries Portfolio tests Physiological measures	Theater tests Rough tests Concept tests	Readability tests Comprehension and reaction tests
Field Methods		
Dummy advertising vehicles	On-air tests	Online theater testing
Posttests		
Field Methods		
Recall tests Association measures	Single-source systems Inquiry tests	Recognition tests Tracking studies

Pretesting Pretests may occur at a number of points, from as early on as idea generation to rough execution to testing the final version before implementing it. More than one type of pretest may be used. For example, concept testing (which is discussed later in this chapter) may take place at the earliest development of the ad or commercial, when little more than an idea, basic concept, or positioning statement is under consideration. In other instances, layouts of the ad campaign that include headlines, some body copy, and rough illustrations are used. For TV commercials, storyboards and animatics may be tested. In these tests specific shortcomings may be identified, and changes made to enhance certain executional elements. As noted by Cramphorn, the best reason to pretest is to identify winners, to enhance good ads, and to eliminate bad ones. He notes that it is important to know the probable effect the ad will have before committing to its use.[7]

The methodologies employed to conduct pretests vary. In focus groups, participants freely discuss the meanings they get from the ads, consider the relative advantages of alternatives, and even suggest improvements or additional themes. In addition to or instead of

page 602

the focus groups, consumers are asked to evaluate the ad on a series of rating scales. (Different agencies use different measures.) In-home interviews, mall intercept, Internet surveys, or laboratory methods may be used to gather the data.

The advantage of pretesting at this stage is that feedback is relatively inexpensive. Any problems with the concept or the way it is to be delivered are identified before large amounts of money are spent in development. Sometimes more than one version of the ad is evaluated to determine which is most likely to be effective.

Because it costs so much less to find out that an ad may not work prior to making it public, rather than after doing so, it certainly makes sense to pretest.

The disadvantage is that mock-ups, storyboards, or animatics may not communicate nearly as effectively as the final product. The mood-enhancing and/or emotional aspects of the message are very difficult to communicate in this format. Another disadvantage is time delays. Many marketers believe being first in the market offers them a distinct advantage over competitors, so they forgo research to save time and ensure this position—even though this may be a risky strategy.

Posttesting Posttesting is also common among both advertisers and ad agencies (with the exception of testing commercials for wearout). Posttesting is designed to (1) determine if the campaign is accomplishing the objectives sought and (2) serve as input into the next period's situation analysis.

Where to Test

In addition to when to test, decisions must be made as to *where.* These tests may take place in either laboratory or field settings.

Laboratory In **laboratory tests**, people are brought to a particular location where they are shown ads and/or commercials. The testers either ask questions about them or measure participants' responses by other methods—for example, pupil dilation, eye tracking, biosciences, or neuroscience measures.

The major advantage of the lab setting is the *control* it affords the researcher. Changes in copy, illustrations, formats, colors, and the like can be manipulated inexpensively, and the differential impact of each assessed. This makes it much easier for the researcher to isolate the contribution of each factor.

The major disadvantage is the lack of *realism.* Perhaps the greatest effect of this lack of realism is a **testing bias**. When people are brought into a lab (even if it has been designed to look like a living room), they page 603 may scrutinize the ads much more closely than they would at home. A second problem with this lack of realism is that it cannot duplicate the natural viewing situation, complete with the distractions or comforts of home. Looking at ads in a lab setting may not be the same as viewing at home on the couch, with the spouse, kids, dog, cat, and parakeet chirping in the background. (A bit later you will see that some testing techniques have made progress in correcting this deficiency. No, they did not bring in the dogs and the parakeets.) Overall, however, the control offered by this method probably outweighs the disadvantages, which accounts for the frequent use of lab methods.

Field Tests **Field tests** are tests of the ad or commercial under natural viewing situations, complete with the realism of noise, distractions, and the comforts of home. Field tests take into account the effects of repetition, program content, and even the presence of competitive messages.

The major disadvantage of field tests is the lack of control. It may be impossible to isolate causes of viewers' evaluations. If atypical events occur during the test, they may bias the results. Competitors may attempt to sabotage the research. And field tests usually take more time and money to conduct, so the results are not available to be acted on quickly. Thus, realism is gained at the expense of other important factors. It is up to the researcher to determine which trade-offs to make.

How to Test

Our discussion of what should be tested, when, and where was general and designed to establish a basic understanding of the overall process as well as some key terms. In this section, we discuss more specifically some of the methods commonly used at each stage. First, however, it is important to establish some criteria by which to judge ads and commercials.

Conducting evaluative research is not easy. Twenty-one of the largest U.S. ad agencies have endorsed a set of principles aimed at "improving the research used in preparing and testing ads, providing a better creative product for clients, and controlling the cost of TV commercials."[8] This set of nine principles, called **Positioning Advertising Copy Testing (PACT)**, defines *copy testing* as research "that is undertaken when a decision is to be made about whether advertising should run in the marketplace. Whether this stage utilizes a single test or a combination of tests, its purpose is to aid in the judgment of specific advertising executions."[9] The nine principles of good copy testing are shown in Figure 18–3.

IGURE 18–3

ositioning Advertising Copy Testing (PACT)

1. Provide measurements that are relevant to the objectives of the advertising.
2. Require agreement about how the results will be used in advance of each specific test.
3. Provide multiple measurements (because single measurements are not adequate to assess ad performance).
4. Be based on a model of human response to communications—the reception of a stimulus, the comprehension of the stimulus, and the response to the stimulus.
5. Allow for consideration of whether the advertising stimulus should be exposed more than once.

6. Require that the more finished a piece of copy is, the more soundly it can be evaluated and require, as a minimum, that alternative executions be tested in the same degree of finish.
7. Provide controls to avoid the biasing effects of the exposure context.
8. Take into account basic considerations of sample definition.
9. Demonstrate reliability and validity.

On the digital side, a blue ribbon task force has been assembled in an industry-wide initiative dedicated to making digital media advertising friendlier to brands by evolving the way interactive advertising is measured. The project, titled "Making Measurement Make Sense," a joint effort of the 4As, the ANA, and the Interactive Advertising Bureau, has a goal of "making digital media measurements directly comparable to those of traditional media, while maintaining the ability to evaluate the unique value that interactivity brings to brand campaigns."[10] The committee has established five guiding principles for this purpose (Figure 18–4).

IGURE 18–4

The Five Guiding Principles of Digital Measurement

Principle #1	Move to a "viewable impressions" standard and count real exposures online.
Principle #2	Online advertising must migrate to a currency based on audience impressions, not gross impressions.
Principle #3	Because all ad units are not created equal, we must create a transparent classification system.
Principle #4	Determine interactivity "metrics that matter" for brand marketers, so that marketers can better evaluate the online's contribution to brand building.
Principle #5	Digital media measurement must become increasingly comparable and integrated with other media.

ource: Author created from www.measurementnow.net.

As you can see, advertisers and their clients are concerned about developing *appropriate* testing methods. Adherence to these principles may not make for perfect testing, but it goes a long way toward improving the state of the art and alleviates at least one of the testing problems cited earlier.

THE TESTING PROCESS

Testing may occur at various points throughout the development of an ad or a campaign: (1) concept generation research; (2) rough, prefinished art, copy, and/or commercial testing; (3) finished art or commercial pretesting; and (4) market testing of ads or commercials (posttesting).

Concept Generation and Testing

Figure 18–5 describes the process involved in advertising **concept testing**, which is conducted very early in the campaign development process in order to explore the targeted consumer's response to a page 605 potential ad or campaign or have the consumer evaluate advertising alternatives. Positioning statements, copy, headlines, and/or illustrations may all be under scrutiny. The material to be evaluated may be just a headline or a rough sketch of the ad. The colors used, typeface, package designs, and even point-of-purchase materials may be evaluated.

IGURE 18–5

Concept Testing

Objective
Explores consumers' responses to various ad concepts as expressed in words, pictures, or symbols.
Method
Alternative concepts are exposed to consumers who match the

characteristics of the target audience. Reactions and evaluations of each are sought through a variety of methods, including focus groups, direct questioning, and survey completion. Sample sizes vary depending on the number of concepts to be presented and the consensus of responses.

Output
Qualitative and/or quantitative data evaluating and comparing alternative concepts.

One of the more commonly used methods for concept testing is focus groups, which usually consist of 8 to 10 people in the target market for the product. Companies have tested everything from product concepts to advertising concepts using focus groups. For most companies, the focus group is the first step in the research process. The number of focus groups used varies depending on group consensus, strength of response, and/or the degree to which participants like or dislike the concepts. Some companies use 50 or more groups to develop a campaign, although fewer than 10 are usually needed to test a concept sufficiently.

While focus groups continue to be a favorite of marketers, they are often overused. The methodology is attractive in that results are easily obtained, directly observable, and immediate. A variety of issues can be examined, and consumers are free to go into depth in areas they consider important. Also, focus groups don't require quantitative analysis. Unfortunately, many managers are uncertain about research methods that require statistics; and focus groups, being qualitative in nature, don't demand much skill in interpretation. Weaknesses with focus groups are shown in Figure 18–6. Clearly, there are appropriate and inappropriate circumstances for employing this methodology.

IGURE 18–6

Veaknesses Associated with Focus Group Research

- The results are not quantifiable.
- Sample sizes are too small to generalize to larger populations.

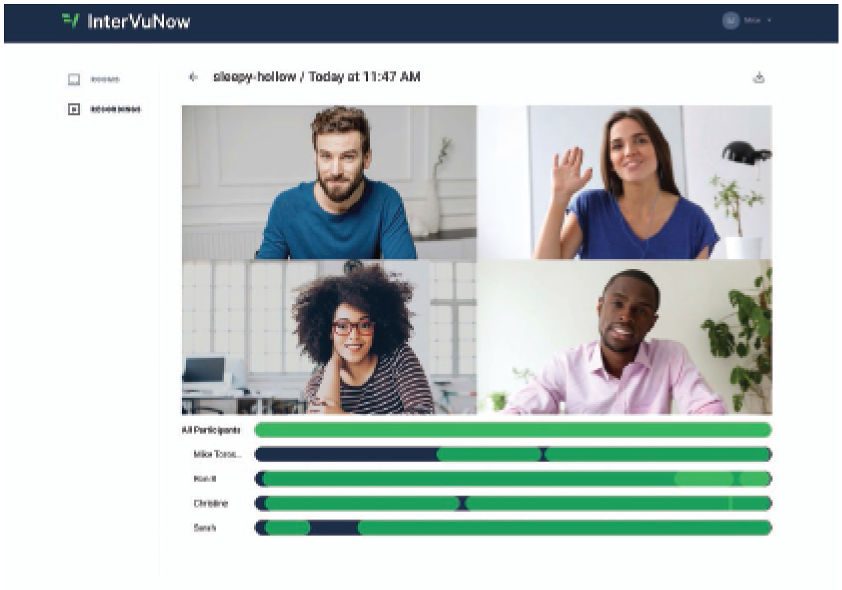
- Group influences may bias participants' responses.
- One or two members of the group may steer the conversation or dominate the discussion.
- Consumers become instant "experts."
- Members may not represent the target market. (Are focus group participants a certain type of person?)
- Results may be taken to be more representative and/or definitive than they really are.

Another way to gather consumers' opinions of concepts is mall intercepts, where consumers in shopping malls are approached and asked to evaluate rough ads and/or copy. Rather than participating in a group discussion, individuals assess the ads via questionnaires, rating scales, and/or rankings. New technologies allow for concept testing over the Internet, where advertisers can show concepts simultaneously to consumers throughout the United States, garnering feedback and analyzing the results almost instantaneously. In addition, online focus groups are being used more often to do concept testing (Exhibit 18–2). Internet methods are becoming increasingly popular given the cost savings and time efficiencies associated with these research methods, and some research firms now offer hybrid studies that involve a combination of online and traditional measures.

XHIBIT 18–2

Online focus groups are used for concept testing.

ource: FocusVision



Rough Art, Copy, and Commercial Testing

Because of the high cost associated with the production of an ad or commercial (many network commercials cost hundreds of thousands of dollars to produce), advertisers are increasingly spending more monies testing a rendering of the final ad at early stages. Slides of the page 606 artwork posted on a screen or animatic and photomatic roughs may be used to test at this stage. (See Figure 18–7 for an explanation of terminology.) Because such tests can be conducted inexpensively, research at this stage is becoming ever more popular.

IGURE 18–7

Rough Testing Terminology

A rough commercial is an unfinished execution that may fall into three broad categories:

Animatic Rough	Photomatic Rough	Live-Action Rough
Succession of drawings/cartoons Rendered artwork Still frames	Succession of photographs Real people/scenery Still frames	Live motion Stand-in/nonunion talent Nonunion crew

Simulated movement: Panning/zooming of frame/rapid sequence	Simulated movements: Panning/zooming of frame/rapid sequence	Limited props/minimal opticals Location settings
Finished Commercial Uses		
Live motion/animation Highly paid union talent Full union crew Exotic props/studio sets/special effects		

But cost is only one factor. The test is of little value if it does not provide relevant, accurate information. Rough tests must indicate how the finished commercial would perform. Studies have demonstrated that these testing methods are reliable and the results typically correlate well with the finished ad.

Most of the tests conducted at the rough stage involve lab settings, although some on-air field tests are also available. Popular tests include comprehension and reaction tests and consumer juries. Again, the Internet allows field settings to be employed at this stage.

1. *Comprehension and reaction tests.* One key concern for the advertiser is whether the ad or commercial conveys the meaning intended. The second concern is the reaction the ad generates. Obviously, the advertiser does not want an ad that evokes a negative reaction or offends someone. **Comprehension and reaction tests** are designed to assess these responses (which makes you wonder why some ads are ever brought to the marketplace).

   Tests of comprehension and reaction employ no one standard procedure. Personal interviews, group interviews, and focus groups have all been used for this purpose, as have online studies, and sample sizes vary according to the needs of the client; they typically range from 50 to 200 respondents.
2. *Consumer juries.* This method uses consumers representative of the target market to evaluate the probable success of an ad. **Consumer juries** may be asked to rate a selection of layouts or copy versions

presented in pasteups on separate sheets. The objectives sought and methods employed in consumer juries are shown in Figure 18–8.

FIGURE 18–8

Consumer Juries

Objective
Potential viewers (consumers) are asked to evaluate ads and give their reactions to and evaluation of them. When two or more ads are tested, viewers are usually asked to rate or rank order the ads according to their preferences.
Method
Respondents are asked to view ads and rate them according to either (1) the order of merit method or (2) the paired comparison method. In the former, the respondent is asked to view the ads and then rank them from one to *n* according to their perceived merit. In the latter, ads are compared only two at a time. Each ad is compared to every other ad in the group, and the winner is listed. The best ad is that which wins the most times. Consumer juries typically employ 50 to 100 participants.
Output
An overall reaction to each ad under construction as well as a rank ordering of the ads based on the viewers' perceptions.

While the jury method offers the advantages of control and cost effectiveness, serious flaws in the methodology limit its usefulness:

- *The consumer may become a self-appointed expert.* One of the benefits sought from the jury method is the objectivity and involvement in the product or service that the targeted consumer can bring to the evaluation process. Sometimes, however, knowing they are being asked to critique ads, participants try to become more *expert* in their evaluations, paying more attention and being more critical than usual. The result may be a less than objective evaluation or an evaluation on elements other than those intended.
- *The number of ads that can be evaluated is limited.* Whether *order of merit* or *paired comparison* methods are used, the ranking procedure becomes tedious as the number of alternatives increases. Consider the ranking of 10 ads. While the top

2 and the bottom 2 may very well reveal differences, those ranked in the middle may not yield much useful information.

In the paired comparison method, the number of evaluations required is calculated by the formula

$$\frac{n(n-1)}{2}$$

If six alternatives are considered, 15 evaluations must be made. As the number of ads increases, the task becomes even more unmanageable.

- *A halo effect is possible.* Sometimes participants rate an ad good on all characteristics because they like a few and overlook specific weaknesses. This tendency, called the **halo effect**, distorts the ratings and defeats the ability to control for specific components. (Of course, the reverse may also occur—rating an ad bad overall due to only a few bad attributes.)
- *Preferences for specific types of advertising may overshadow objectivity.* Ads that involve emotions or pictures may receive higher ratings or rankings than those employing copy, facts, and/or rational criteria. Even though the latter are often more effective in the marketplace, they may be judged less favorably by jurists who prefer emotional appeals.

Some of the problems noted here can be remedied by the use of ratings scales instead of rankings. But ratings are not always valid either. Thus, while consumer juries have been used for years, questions of bias have led researchers to doubt their validity. As a result, a variety of other methods (discussed later in this chapter) are more commonly employed.

A/B Testing The process of **A/B testing** has been employed by marketers for years. The process involves the testing of two versions of an advertisement or homepage to see which will be the more effective prior to launch. In the vast majority (if not all) of these cases the ads or web pages are finished products. In making a case for using A/B testing earlier in the development process, Molly Soat makes the argument that waiting so long to test could be costing companies money, and at the same time be less effective. She notes that a number of advertising experts call for the testing of display art, or scripts, for example to gain insights into consumers' thinking rather than waiting to just have them choose between two finished ads. Besides saving a lot of money by not developing and producing a page 608 potentially ineffective ad, the results would likely lead to ads that use consumers' insights. The experts argue for using A/B testing earlier in the developmental process and measuring consumer's emotional and neurological responses rather than just have them choose between two finish products. Of course, the finished ads could be A/B tested as well.[11]

Pretesting of Finished Ads

Pretesting finished ads is one of the more commonly employed studies among marketing researchers and their agencies. At this stage, a finished advertisement or commercial is used; since it has not been presented to the market, changes can still be made.

Many researchers believe testing the ad in final form provides better information. Several test procedures are available for print and broadcast ads, including both laboratory and field methodologies.

Print methods include portfolio tests, analyses of readability, and dummy advertising vehicles. Broadcast tests include theater tests and on-air tests. Both print and broadcast may use physiological measures.

Pretesting Finished Print Messages A number of methods for pretesting finished print ads are available. One is *Gallup & Robinson's Impact System*, described in Figure 18–9. The most common of these methods are portfolio tests, readability tests, and hybrid measures.

IGURE 18–9

G&R Research and Consulting

Objective
Understanding the performance of individual advertising executions or testing finished products. Research and testing gauge the potential of an idea behind potential advertising executions prior to fielding.

Methods
In-person interviewing; telephone; Internet; physiological; advertising response modeling; qualitative and quantitative research; facial EMG; others.
Output
Sales-validated methods for the development of messaging and the optimization of performance.

Portfolio Tests Portfolio tests are a laboratory methodology designed to expose a group of respondents to a portfolio consisting of both control and test ads. Respondents are then asked what information they recall from the ads. The assumption is that the ads that yield the *highest recall* are the most effective.

While portfolio tests offer the opportunity to compare alternative ads directly, a number of weaknesses limit their applicability:

1. Factors other than advertising creativity and/or presentation may affect recall. Interest in the product or product category, the fact that respondents know they are participating in a test, or interviewer instructions (among others) may account for more differences than the ad itself.
2. Recall may not be the best test. Some researchers argue that for certain types of products (those of low involvement) ability to recognize the ad when shown may be a better measure than recall.

One way to determine the validity of the portfolio method is to correlate its results with readership scores once the ad is placed in the field. Whether such validity tests are being conducted or not is not readily known, although the portfolio method remains popular in the industry. A variety of the portfolio test is the **mock magazine test** in which an ad is placed in an actual magazine and a similar methodology is employed.

Readability Tests The communications efficiency of the copy in a print ad can be tested without reader interviews. (Note: The tests are now also being used to test the readability of websites.) This test uses the **Flesch formula**, named after its developer, Rudolph Flesch, to assess readability of the copy by determining the average number of syllables per 100 words. Human interest appeal of the material, length of sentences, and familiarity with certain words are also considered and correlated with the educational background of target audiences. Test results are compared to previously established norms for various target audiences. The test suggests that copy is best comprehended when sentences are short, words are concrete and familiar, and personal references are drawn.

The Flesch Kincaid Reading Ease Score method eliminates many of the interviewee biases associated with other tests and avoids gross errors in understanding. The norms offer an attractive standard for comparison. Other readibility measures also exist, including the SMOG Readability Formula and the Fry Graph Readability Formula.

Disadvantages are also inherent, however. The copy may become too mechanical, and direct input from the receiver is not available. Without this input, contributing elements like creativity cannot be addressed. To be effective, this test should be used only in conjunction with other pretesting methods.

New Print Pretesting Measures In an effort to improve upon the traditional print pretest measures, a number of companies have introduced new methodologies or improved upon existing methods that employed dummy advertising vehicles. Many of these involve hybrid measures that either measure effectiveness in different ways or in different channels. For example, Ipsos-ASI Next*Connect has introduced an online copy testing tool (Figure 18–10) that can test ads on digital or traditional media. Consumers are recruited to complete an online survey where they are exposed to a variety of ad messages. The results are compared against a control group that was not exposed to the advertising. The methodology allows for testing ads in finished or rough formats, for individual executions or multiple campaign elements, and determine ads’ impact in traditional and new media (Exhibit 18–3).

IGURE 18–10

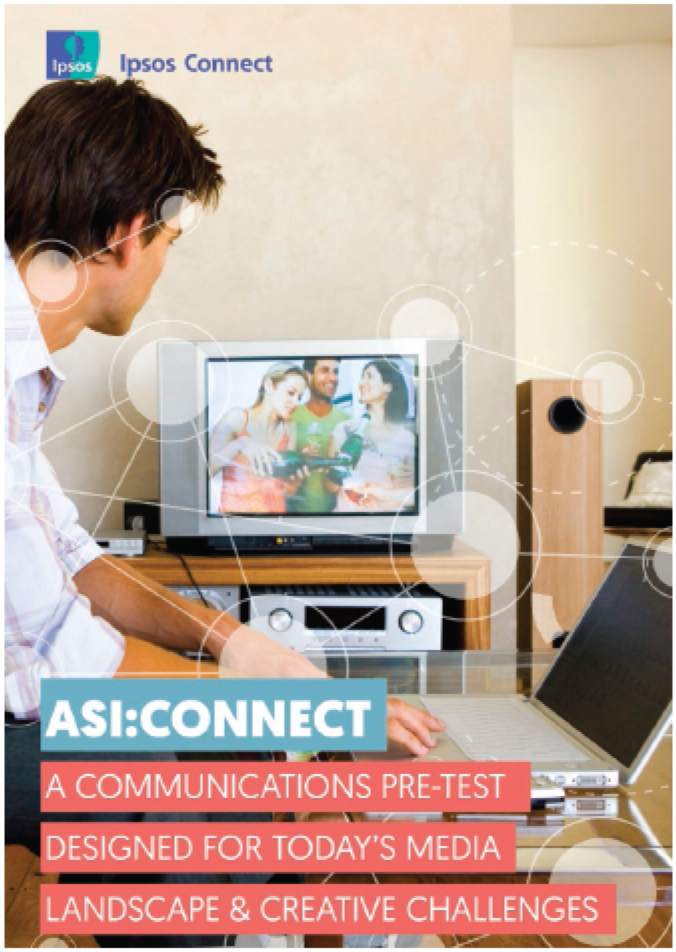
psos-ASI's Next*Connect

Objective
To assist advertisers in copy testing of rough or finished advertisements in traditional or digital form to determine the individual executions or multiple campaign elements across any platform.
Method
Consumers are recruited to complete an online survey where they are exposed to a variety of messages and results are compared against those from a control group not exposed to any ads.
Output
Standard scores, related recall, persuasion, and comprehensive diagnostics; potential ad-driven in-market sales.

XHIBIT 18–3

psos-ASI offers a comprehensive testing measure.

ource: Ipsos



The PTG (PreTesting Group) methodology measures the time a respondent spends with a print ad or tablet or with a hidden camera in a store. The altered (dummy) magazines contain a number of ads targeted toward the interests of the consumers. Unknown to the consumers, as they go through the magazines, hidden cameras using eye-tracking technology record where their eyes go on the page, how long they stay on the page, and the stopping power of the ad. The studies can be conducted in person or online and can measure recall, main idea communication, and more.

page 610

Kantar Millward Brown offers its own trademarked methodology as well, called Link. According to the company, Link uses a comprehensive set of diagnostic questions to evoke viewer reactions to the ads. Nonverbal measures including eye tracking can also be used to determine consumers' enjoyment, comprehension, involvement, and other reactions to the ads. The three key metrics provided include awareness, persuasion, and short-term sales likelihood.

While other methods and measures are available, the most popular form of pretesting of print ads now involves a series of measures. The tests can typically be used for rough and/or finished ads and are most commonly conducted in a lab or online.

Pretesting Finished Broadcast Ads A variety of methods for pretesting broadcast ads are available. The tests can typically be used for rough and/or finished ads and are most commonly conducted in a lab or online. As noted, most of the companies that offer pretesting of print ads have expanded their services to include TV and other platforms. (The ASI method described in Figure 18–10 uses the same testing platform across media.) For example, MSW•ARS Research provides offerings that allow the advertiser to evaluate the impact of advertising messages and campaigns composed of any combination of touch points, including broadcast (television and radio), outdoor, and digital. The company provides research solutions in brand strategy, all stages of creative development from concept to fully finished ads, campaign evaluation across all channels, advertising and brand equity tracking, and additional consulting services.

Theater Tests In the past, one of the most popular laboratory methods for pretesting finished commercials was **theater testing**. In theater tests participants are invited to view pilots of proposed TV programs. In some instances, the show is actually being tested, but more commonly a standard program is used so audience responses can be compared with normative responses established by previous viewers. Sample sizes range from 250 to 600 participants, with 300 being most typical.

The methods of theater testing operations vary, though all measure brand preference changes. For example, many of the services now use programs with the commercials embedded for viewing in one's home or office rather than in a theater. Others establish viewing rooms in malls and/or hotel conference rooms. Some do not take all the measures listed here; others ask the consumers to turn dials or push buttons on a keypad to provide the continual responses. An example of one methodology is shown in Figure 18–11 (you may remember GfK from Chapter 10) and Exhibit 18–4.

IGURE 18–11

GfK AD*VANTAGE Pulse Methodology

AD*VANTAGE Pulse surveys are based on tests carried out in a studio. The number of respondents is typically about *n* = 125, normally composed of category users defined by the client. Each participant is shown a TV program, with commercial breaks that include the test commercial, and then asked questions about the program. The closed- and open-ended questions are recorded via a touch-screen system. Key measures include (1) visibility (Will the commercial be remembered?); (2) branding (Will the brand be remembered?); (3) communications (What visuals and messages will be remembered?); (4) brand enhancement (Does the ad promote a positive feeling toward the brand?); and (5) persuasion (Will the commercial inspire nonusers to try the product, and will it enhance brand loyalty among existing customers?). The diagnostics include measures of viewers' awareness, comprehension, uniqueness, and involvement of the commercial. The AD*VANTAGE Pulse methodology also allows for scene-by-scene analysis, and can be used to test all traditional forms of advertising as well as digital.

XHIBIT 18–4

This ad for GfK promotes its research and testing as a smart business decision. Do you agree or disagree?

ource: GfK

Those opposed to theater tests cite a number of disadvantages. First, they say the environment is too artificial. The recruiting and lab testing conditions are likely to lead to testing effects, and consumers are exposed to commercials they might not even notice in a more natural environment. Second, the contrived measure of brand preference change seems too phony to believe. Critics contend that participants will see through it and make changes just because they think they are supposed to. Finally, the group effect of having others present and overtly exhibiting their reactions may influence viewers who did not have any reactions themselves.

Proponents argue that theater tests offer distinct advantages. In addition to control, the established norms (averages of commercials' performances) indicate how one's commercial will fare against others in the same product class that were already tested. Further, advocates say the brand preference measure is supported by actual sales results.

Despite the limitations of theater testing, most major consumer-product companies have used it to evaluate their commercials. This method may have shortcomings, but it allows them to identify strong or weak commercials and to compare them to other ads. Companies like Ipsos-ASI and Consumer Quest now offer in-home and online theater testing (Exhibit 18–4).

On-Air Tests Some of the firms conducting theater tests also insert the commercials into actual TV programs in certain test markets. Typically, the commercials are in finished form, although the testing of ads earlier in the developmental process is becoming more common. This is referred to as an **on-air test** and often includes single-source ad research (discussed later in this chapter). Information Resources, Ipsos-ASI, MSW•ARS, and Nielsen are well-known providers of on-air tests.

On-air testing techniques offer all the advantages of field methodologies, as well as all the disadvantages. The most commonly employed metric used in an on-air test is **recall**—that is, the number of persons able to recall the ad and/or its message. In an examination of real-world advertising tests reported in the *Journal of Advertising Research,* Hu et al. conclude that recall and persuasion pretests, while often employed, do not fare well in respect to reliability and/or validity.[12] Nevertheless, most of the testing services have offered evidence of both validity and reliability for on-air pretesting of commercials. Both Ipsos-ASI and MSW•ARS claim their pretest and posttest results yield the same recall scores 9 out of 10 times—a strong indication of reliability and a good predictor of the effect the ad is likely to have when shown to the population as a whole.

In summary, on-air pretesting of finished or rough commercials offers some distinct advantages over lab methods and some indications of the ad's likely success. Whether the measures used are as strong an indication as the providers say remains in question.

Physiological Measures A less common but increasingly adopted method of pretesting finished commercials involves a laboratory setting in which physiological responses are measured. These measures indicate the receiver's *involuntary* response to the ad, theoretically eliminating biases associated with the voluntary measures reviewed to this point. (Involuntary responses are those over which the individual has no control, such as heartbeat and reflexes.) Physiological measures used to test both print and broadcast ads include pupil dilation, galvanic skin response, eye tracking, and brain waves:

1. *Pupil dilation.* Research in **pupillometrics** is designed to measure dilation and constriction of the pupils of the eyes in response to stimuli. Dilation is associated with action; constriction involves the body's conservation of energy.

   Advertisers have used pupillometrics to evaluate product and package design as well as to test ads. Pupil dilation suggests a stronger interest in (or preference for) an ad or implies arousal or attention-getting capabilities. Other attempts to determine the affective (liking or disliking) responses created by ads have met with less success.

   Because of high costs and some methodological problems, the use of pupillometrics has waned over the past decade. But it can be useful in evaluating certain aspects of advertising.
2. *Galvanic skin response (GSR).* Also known as **electrodermal response (EDR)**, GSR measures the skin's resistance or conductance to a small amount of current passed between two electrodes. page 612 Response to a stimulus activates sweat glands, which in turn increases the conductance of the electrical current. Thus, GSR/EDR activity might reflect a reaction to advertising. While there is evidence that GSR/EDR may be useful to determine the effectiveness of ads, difficulties associated with this testing method have resulted in its infrequent use at this time.
3. *Eye tracking.* As seen in Digital and Social Media Perspective 18–2 a methodology that is more commonly employed is **eye tracking** (Figure 18–12), in which viewers are asked to view an ad while a page 613 sensor aims a beam of infrared light at the eye. The beam follows the movement of the eye and shows the exact spot on which the viewer is focusing, and for how long (Exhibit 18–5). The continuous reading of responses demonstrates which elements of the ad are attracting attention, how long the viewer is focusing on them, and the sequence in which they are being viewed.

FIGURE 18–12

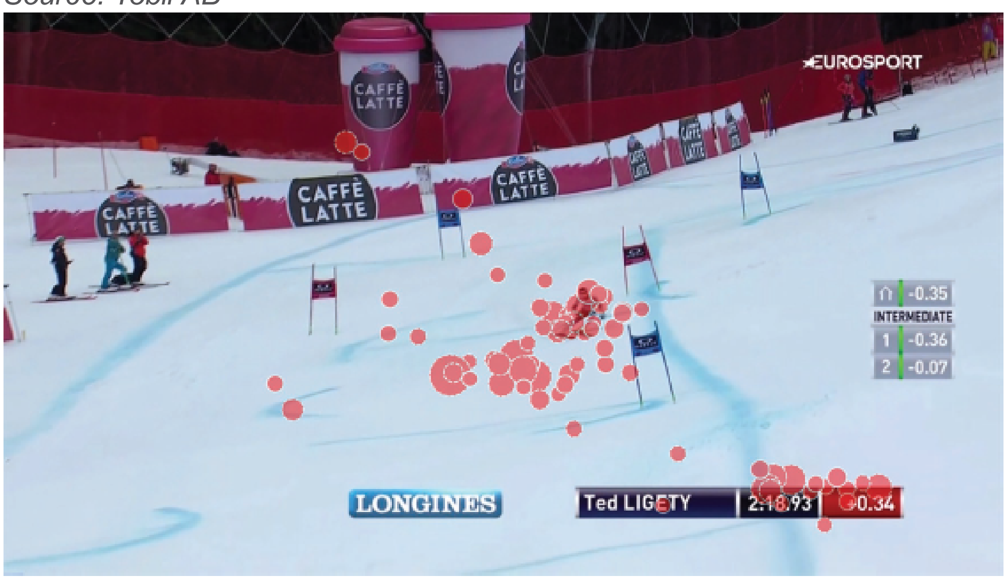
Eye Movement Research

Objective
Track viewers' eye movements to determine what viewers read or view in print ads and where their attention is focused in TV commercials, at websites, or on billboards.
Method
Fiber optics, digital data processing, and advanced electronics are used to follow eye movements of viewers and/or readers as they process an ad.
Output
Relationship among what readers see, recall, and comprehend. Scan movement paths in print ads, on billboards, in commercials, in print materials, and on websites. (Can also be used to evaluate package designs.)

EXHIBIT 18–5

SensoMotoric Instruments offers eye-tracking solutions as it becomes more commonly used in digital testing.

Source: Tobii AB



Eye tracking can identify strengths and weaknesses in an ad. For example, attractive models or background action may distract the viewer's attention away from the brand or product being advertised. The advertiser can remedy this distraction before fielding the ad. In other instances, colors or illustrations may attract attention and create viewer interest in the ad.

Eye tracking has increasingly been used to measure the effectiveness of websites and online ads and, as noted earlier, e-mails. Using eye tracking to examine how consumers view homepages, Steve Outing and Laura Roel were able to determine that (1) eyes first fixate on the upper left of the screen, (2) dominant headlines draw attention first, and (3) larger type promotes scanning, while small type encourages reading. The study drew other conclusions as well—too many to mention here.[13] Digital and Social Media Perspective 18–2 details some other applications of eye tracking and other physiological methods.

4. *Brain waves.* **Electroencephalographic (EEG) measures** can be taken from the skull to determine electrical frequencies in the brain. These electrical impulses are used in two areas of research, alpha waves and hemispheric lateralization:

   - **Alpha activity** refers to the degree of brain activation. People are in an alpha state when they are inactive, resting, or sleeping. The theory is that a person in an alpha state is less likely to be processing information (recall correlates negatively with alpha levels) and that attention and processing require moving from this state. By measuring a subject's alpha level while viewing a commercial, researchers can assess the degree to which attention and processing are likely to occur.
   - **Hemispheric lateralization** distinguishes between alpha activity in the left and right sides of the brain. It has been hypothesized that the right side of the brain processes visual stimuli and the left processes verbal stimuli. The right hemisphere is thought to respond more to emotional stimuli, while the left responds to logic. page 615 The right determines recognition, while the left is responsible for recall. If these hypotheses are correct, advertisers could design ads to increase learning and memory by creating stimuli to appeal to each hemisphere. However, some researchers believe the brain does not function laterally and an ad cannot be designed to appeal to one side or the other.

- Using technologies originally designed for the medical field such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG), neuroscientists have teamed up with marketers to examine physiological reactions to ads and brands through brain scan imaging. By monitoring the brain activity directly, scientists are learning how consumers make up their minds by measuring chemical activity and/or changes in the magnetic fields of the brain as well as how they react to commercials.

Digital and Social Media Perspective 18–2 > > >

Academics and Practitioners Agree That Physiological Measures of Effectiveness Work

An increased emphasis on accountability and the pressure to determine an ad's effectiveness, along with the explosion of new media and changes and improvements to existing media, are just some of the reasons why marketers continue to seek new ways to examine how their messages are performing. As a result, new research methodologies are being developed, and existing technologies are being adapted to assist in making this determination. Many of these involve the application of neurosciences. Recognizing the demand for greater efficiency and effectiveness, Heather Andrew of Neuro-Insight published a paper on how neuroscience principles can help deliver more effective brand communications. Andrew notes that "by focusing on the subconscious drivers of behavior, neuroscience offers a way of understanding the link between creative and media choices and the ultimate effectiveness of a campaign." The reason, she says, is that in this way marketers are able to look beyond consumers' claimed motivations to understand what is really going on in people's heads. Some of these are proving to be quite successful:

On the academic side, Innerscope and professors at Temple University joined together to measure physiological reactions to Super Bowl ads including heart rates and breathing patterns as well as fMRI data. Another team of marketing professors used visual-imagery theory to test whether consumers were making the connection between the sponsor and the event they were attending—their results indicated that "attendees who rated the event as 'higher quality' had a higher attitude toward the sponsors' products that were showcased at the tournament."

Practitioners have also joined in. As many advertisers know that click-through measures are not the best measures of ads on the web, another company, RealVu, measures the seconds a user is actually exposed to an ad. While companies agree that click-throughs are not effective measures of Internet advertising, they don't necessarily agree as to what is the best measure.

A methodology that seems to be gaining more and more acceptance in marketing circles has actually been around for quite some time. The usefulness of eye-tracking technology for measuring advertising effectiveness is not news for marketers. Back in the late 1970s and early 1980s, advertising researchers were employing this technology to determine where viewers were focusing when they looked at ads and/or TV commercials. One story is that one of the original Tab diet soft-drink commercials was changed when viewers focused on the female swimsuit model instead of the product. The commercial was supposedly changed to have her hold the can in front of her body to attract attention. For whatever reason, while proven to be useful, eye tracking never went away, but no one seemed to pay much attention to it. Now all of that is changing.

One company heavily involved with using eye tracking to determine advertising effectiveness is Tobii. An international company with offices located in the United States, Norway, Germany, Japan, and China, Tobii was founded as a result of a research project at the Royal Institute of Technology in Stockholm, Sweden, and employs eye tracking for a variety of nonmarketing purposes, as well as for measuring ad effectiveness. The company offers its research services for studying online marketing, including determining the effectiveness of banner ads, e-mail campaigns, newsletters, and in-game advertising. More traditional testing includes copy testing print ads in the creative stage; assessing the impact of out-of-home ads in bars, airports, grocery stores, and so on; determining the effectiveness of product placements and TV commercials; and even testing promotional products.

There are a variety of reasons marketers have found a number of new applications for eye tracking. In addition to the effectiveness measures offered by Tobii and others, companies have used this methodology for:

- *Retail (interior displays/window design):* Eye-tracking studies enabling interior designers to choose color, lighting, architecture, and so on to subtly influence the customer's mood. This includes the use of eye catchers to attract passers-by and arouse their curiosity.
- *Package design:* Optimizing the look and feel of the packaging as well as the attention-getting capabilities.
- *Advertisements:* Allowing the advertiser to determine what customers see and to determine the effectiveness of copy and visuals.

- *Online marketing:* From web design to viewers' attention to display ads, marketers are finding eye tracking a valuable tool to enhance effectiveness.

Researchers use eye movements to tell specifically where viewers are looking, to see if they are reading or scanning, and what catches their first view. They can also tell how much time the viewers spend looking at a specific stimulus and the order in which they view others (if they do!). For example:

Mark Nazh/Shutterstock

- An eye-tracking study conducted by Oneupweb asked participants to navigate Facebook, Twitter, and YouTube as they normally would. The study revealed that 65 percent of them engaged with a sponsor within 10 seconds of beginning their search, when asked to search for Pepsi on both Facebook and YouTube (Twitter was not included because it has no search tool). On Facebook, the results showed that their first attention was paid to the sponsored ads, and others were barely, if at all, viewed, while on YouTube, the top six organic results and the first sponsored ad got all the attention. The conclusion? Sponsored ads work better than expected.
- In another Internet study, Jakob Nielsen and Kara Pernice, authors of the book *Eyetracking Web Usability,* tracked how people navigate websites when looking for information.

Searches included how to deal with heartburn, shopping for baby presents, and picking cell phone features, among others. Their results indicate that to be effective, an ad should be simple, with those that had only text or text and a separate image being most effective, while those that imposed text on top of images or included animation fared the worse. Although the study showed that people saw 36 percent of the ads on the pages they visited, the time spent viewing them was only one-third of a second. One of the most surprising results was the fact that text-only ads scored best. When asked for an explanation, the authors hypothesized that unlike television which is a passive medium, the Web is all about taking action, searching, clicking, and so forth. On the Web, it is the value of content that makes it attractive.

- In a scarier application, a Canadian company now offers an eye-tracking device that can determine when someone looks at a billboard, where the person looks, and for how long, without one even knowing it—up to 33 feet. It then provides Google-like metrics that greatly improve on existing methods of data collection.
- In Germany, a pre- and posttest research study was designed to measure participants' perceptions and recall of ads through a virtual reality situation in which one drives a van down the street with advertisements on both sides. The eye-tracking and survey results are then combined to determine viewers' unaided/aided recall, recognition and design aspects of the ads. And yes, they can do it at different speeds!

Part of the reason for eye-tracking measures not hitting the mainstream is the high cost and obtrusiveness of the measurement equipment (usually the participants have to wear headgear, whether in a lab setting or in the field—for example, walking through a department store or supermarket looking weird). There was also the question about whether the fact that the participants knew they were in a study influenced their behaviors. However, P&G, working with an external eye-tracking firm, believes that eye tracking has saved the company substantial amounts of money by allowing P&G to cancel website display ads that are not being seen. The information is gathered from a panel of consumers using webcams to track their eye movements while on a site, and is then fed back to P&G's creative department indicating what works best. In his latest book, Harvard Business School Professor Gerald Zaltman provides even more insights into why eye tracking is important. The conclusion? Eye tracking allows researchers to determine if viewers get the message. If they do get it, eye tracking will tell us why. If the viewer doesn't get it the analysis will give us the data needed to determine why not.

Maybe eye tracking is here to stay!

Sources: "Neuroscience Techniques Can Drive Effectiveness," January 11, 2019, www.warc.com; Karna, "Why Is Eye-Tracking Important for Market Research," April 23, 2018; www.medium.com; Angeline G. Close, Russell Lacey, and T. Bettina Cornwell, "Visual Processing and Need for Cognition Can Enhance Event-Sponsorship Outcomes—How Sporting Event Sponsorships Benefit from the Way Attendees Process Them," *Journal of Advertising Research* 55, no. 2 (May 2015), pp. 206–15; Tom Avril, "Super Bowl Ads Get Look by Science," February 27, 2014, www.philly.com; Christopher Heine, "P&G Slashes Wasteful Display Ad Budgets Using Eye Tracking," May 30, 2013, www.adweek.com; Nathania Johnson, "Eye Tracking Study Shows Sponsored Ads Attract Social Media Searchers," July 19, 2009, blog.searchenginewatch.com; Dan Skeen, "Eye-Tracking Device Lets Billboards Know When You Look at Them," June 12, 2007, www.wired.com; G. Theuner, K. Pischke, and T. Bley, "Analysis of Advertising Effectiveness with Eye Tracking," *Proceedings of Measuring Behavior 2008,* Maastricht, The Netherlands, August 26–29, 2008, pp. 229–30.

EEG research has engaged the attention of academic researchers for some time, but recently the technology has gained in attractiveness to practioners as well. In addition to those examples cited earlier in the chapter and throughout this text, companies like the Nielsen Company—the world's largest audience measurement company—have invested in neuroscience measurement as *Fortune* 500 companies increasingly seek these services. Neuroscience measures have been used to determine viewers' responses to Super Bowl commercials, differences in responses to drinking Coke and Pepsi, political ads, and movie trailers, to name just a few of the many applications.

Market Testing of Ads

The fact that the ad and/or campaign has been implemented does not diminish the need for testing. The pretests were conducted on smaller samples and may in some instances have questionable merit, so the marketer must find out how the ad is doing in the field. In this section, we discuss methods for posttesting an ad. Many of the tests are similar to the pretests discussed in the previous section and are provided by the same companies.

Posttests of Print Ads A variety of print posttests are available, including inquiry tests, recognition tests, and recall tests.

Inquiry Tests Used in both consumer and business-to-business market testing, **inquiry tests** are designed to measure advertising effectiveness on the basis of inquiries generated from ads appearing in various print media, often referred to as "bingo cards" such as the one shown in Exhibit 18–6. While still used, the response card is employed less often today as viewers can seek information merely by searching on the URL provided in the ad. Magazines used in B2B marketing are more likely to use inquiry cards. However, no one seems to know if the leads are followed up on. At least one study credits the cutback on the use of inquiry cards as responsible for declines in B2B magazine circulation.[14] The inquiry may take the form of the number of coupons returned, phone calls generated, or direct inquiries through reader cards. If you called in response to an ad in a local medium recently, perhaps you were asked how you found out about the company or product or where you saw the ad. This is a very simple measure of the ad's or medium's effectiveness.

XHIBIT 18–6

Reader response cards are popular in business-to-business markets.

ource: Target Marketing/NAPCO Media

TARGET MARKETING Receive the information you need to market smarter instantly!
Go to: www.targetonline.com and click on free information.

Or...Fill out this card and either FAX your request today: 215-238-5388 or mail.

Free Subscription

Do you wish to receive (continue to receive) a FREE subscription to TARGET MARKETING? ❑ YES ❑ NO

Signature ____________ Date ______

Name (please print) ____________ Title ______

Company Name ____________

Address ____________

City/State/Zip ____________

Telephone ____________ FAX ____________

Email Address: ____________

1. TYPE OF BUSINESS AT THIS LOCATION: (Check One)
   - 04 ☐ Catalogers, Mail Order Companies
   - 05 ☐ Retailers, Wholesaler/Distributor
   - 06 ☐ On-line Merchant
   - 03 ☐ Financial Services: Banks, Insurance, Credit Cards, Investment
   - 21 ☐ Fund Raisers: Non-Profit Groups, Charities
   - 33 ☐ High Technology: Internet Services, Computers, Software, Telecommunications, Office Equipment
   - 12 ☐ Manufacturers, Package Goods and Other Business-to-Business Marketers
   - 31 ☐ Non Financial Services Marketers: Associations, Clubs, Educational, Government Agencies, Health Care, Membership Organizations, Personnel Services, Real Estate, Utilities, Communications, Tourism and Travel
   - 13 ☐ Publishers: Magazines, Books, Newsletters, Newspapers, Book & Record Clubs On-line publications
   - 01 ☐ Advertising/DM Agencies, Freelance Creative and Consultants
   - 09 ☐ Mailing List Professionals: Brokers, Compilers, Owners
   - 20 ☐ Other: (Please Specify) ____________
2. JOB FUNCTION: (Check One)
   - 10 ☐ Corporate and General Management (Excluding Sales, Marketing and Advertising Management)
   - 02 ☐ Marketing, List and Sales Management
   - 01 ☐ Advertising Promotion and Public Relations Management
   - 08 ☐ Operational, Technical and Telemarketing Management
   - 06 ☐ Circulation Management
   - 11 ☐ Other: (Please Specify) ____________
3. MY FIRM'S MARKETING ACTIVITY: (Check All That Apply)
   - 1 ☐ Catalog
   - 2 ☐ Direct Mail
   - 09 ☐ Internet
   - 3 ☐ Telemarketing
   - 10 ☐ Alternate Media
   - 5 ☐ Consumer
   - 6 ☐ Business-to-Business
   - 7 ☐ International
   - 8 ☐ None of the Above
4. IN THE PERFORMANCE OF YOUR JOB, CHECK THE ONE THAT BEST DESCRIBES YOUR PURCHASING AUTHORITY: (Check One)
   - 1 ☐ Authorize/Approve Purchase
   - 2 ☐ Recommend/ Specify Purchase
   - 3 ☐ No Purchasing Authority

Free Information

Circle the free info number for the products and services that interest you

001	002	003	004	005	006	007	008	009	010	011	012
013	014	015	016	017	018	019	020	021	022	023	024
025	026	027	028	029	030	031	032	033	034	035	036
037	038	039	040	041	042	043	044	045	046	047	048
049	050	051	052	053	054	055	056	057	058	059	060
061	062	063	064	065	066	067	068	069	070	071	072
073	074	075	076	077	078	079	080	081	082	083	084
085	086	087	088	089	090	091	092	093	094	095	096
097	098	099	100	101	102	103	104	105	106	107	108
109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132
133	134	135	136	137	138	139	140	141	142	143	144
145	146	147	148	149	150	151	152	153	154	155	156
157	158	159	160	161	162	163	164	165	166	167	168
169	170	171	172	173	174	175	176	177	178	179	180

For Information on Product/Service Categories

To request information on any of the following areas of interest, check the appropriate free information number:

- 501 ☐ Agencies/Consultants/Creative Services
- 502 ☐ Alternate Media/Delivery
- 503 ☐ Broadcast Media/Services
- 504 ☐ Call Center Equipment
- 505 ☐ Credit/Collection Services
- 506 ☐ Database Management Systems/ Services/Software
- 507 ☐ Delivery Services
- 529 ☐ E-mail Marketing
- 508 ☐ Envelopes/Order Forms/Labels
- 509 ☐ Fulfillment Services/Operations
- 510 ☐ Int'l Call Centers
- 511 ☐ Int'l Delivery Services
- 512 ☐ Int'l DM Agencies
- 525 ☐ International DM Services
- 530 ☐ Internet Services
- 513 ☐ Lettershop Services
- 514 ☐ Lists-Business-to-Business
- 515 ☐ Lists-Consumer
- 528 ☐ Lists-E-mail
- 516 ☐ Lists-International
- 517 ☐ List Services
- 518 ☐ Mailing Equipment/Services
- 520 ☐ Paper
- 521 ☐ Premiums/Promotional Products
- 522 ☐ Prepress Equipment/Services
- 523 ☐ Printers
- 527 ☐ Printing Systems/Equipment
- 524 ☐ Telemarketing Services

RS0110

More complex methods of measuring effectiveness through inquiries may involve (1) running the ad in successive issues of the same medium; (2) running **split-run tests**, in which variations of the ad appear in different copies of the same newspaper or magazine; and/or (3) running the same ad in different media. Each of these methods yields information on different aspects of the strategy. The first measures the *cumulative* effects of the campaign; the second examines specific elements of the ad or variations on it. The final method measures the effectiveness of the medium rather than the ad itself.

While inquiry tests may yield useful information, weaknesses in this methodology limit its effectiveness. For example, inquiries may not be a true measure of the attention-getting or information-providing

page 616

aspects of the ad. The reader may be attracted to an ad, read it, and even store the information but not be motivated to inquire at that particular time. Time constraints, lack of a need for the product or service at the time the ad is run, and other factors may limit the number of inquiries. But receiving a small number of inquiries doesn't mean the ad was not effective; attention, attitude change, awareness, and recall of copy points may all have been achieved. At the other extreme, a person with a particular need for the product may respond to any ad for it, regardless of specific qualities of the ad.

Major advantages of inquiry tests are that they are inexpensive to implement and they provide some feedback with respect to the general effectiveness of the ad or medium used. But they are usually not very effective for comparing different versions or specific creative aspects of an ad.

Recognition Tests Perhaps the most common posttest of print ads is the **recognition method**, most closely associated with GfK Starch. The *Starch Ad Readership Report*, which has existed for over 85 years, lets the advertiser assess the impact of an ad in a single issue of a magazine, over time, and/or across different magazines. Starch measures hundreds of thousands of print ads representing more than 180 consumer magazines per year and provides a number of measures of an ad's effectiveness. In addition to the traditional *Starch Ad Readership Report*, Starch now offers syndicated and private studies, pretesting, out-of-home studies, and more. The measures used in the Starch test methodology are shown in Figure 18–13. An example of a Starch scored ad is shown in Exhibit 18–7.

IGURE 18–13

The GfK Starch Ad Readership Report

Objective
Determining recognition of print ads and providing insight into the involvement readers have with specific ads. Starch now provides this service for ads appearing in magazines and/or digital.
Method

Personal interviewers screen readers for qualifications and determine exposure and readership to specific issues of newspapers, consumer magazines, and business and professional publications. Participants are asked to go through the magazines, looking at the ads, and provide specific responses.

Output

Starch Ad Readership Reports generate four recognition scores:

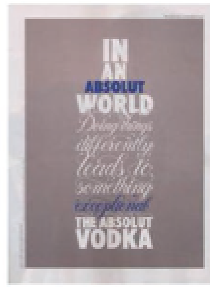
- Noting score—the percentage of readers who remember seeing the ad.
- Brand-associated score—the percentage of readers who recall seeing or reading any part of the ad identifying the product or brand.
- Read any score—the percentage of readers who read any of the ad's copy.
- Read-most score—the percentage of readers who report reading at least half of the copy portion of the ad. Ad norms provide a benchmark to provide a comparison to other ads in the issue.

These conventional scores have now been supplemented by additional measures, including brand disposition, purchase behavior or intention, actions taken, word-of-mouth opportunity, and publication and advertising engagement.

XHIBIT 18–7

An ad scoring very high on the Starch rating scale.

ource: GfK MRI



GfK MRI	Starch® Advertising Research	Entertainment Weekly - September 11, 2009

ABSOLUT VODKA
Category: Other Alcoholic Beverages
1P4| Page 29

SAMPLE Report
For Illustrative Purposes Only

Issue Audience (000): 10,555	Noted	Assoc.	Read Any	Read Most
Advertiser	72%	68%	59%	-
AdMeasure Audience (000)	7,600	7,177	6,227	
Issue Norm (17 Ads)	61%	54%	48%	18%
Issue Index	118	126	123	144
Comparable MRI Starch Adnorm: 1P4 - Other Alcoholic Beverages				
Adnorm (156 Ads)	52%	45%	42%	20%
Adnorm Index	138	151	140	NA

Actions Taken by Those Who Noted the Ad	Any Actions Taken	Purchased the product/ service	Visited their website	Clipped or saved the ad	Contacted company by mail/phone	Looked for the product/ service	Recommended the product to someone	Talked to a doctor about the product (Rx ads only)	Talked to someone about it	None
Advertiser	17%	13%	1%	1%	0%	1%	3%	NA	2%	83%
Category Norm (268 Ads)	43%	12%	6%	4%	1%	12%	8%	NA	15%	57%
Category Index	40	108	17	25	0	8	38	NA	13	146

Brand Disposition	Positively disposed (Net)	My favorite brand	One of several brands I like	Don't use, but it's worth trying	Negatively disposed (Net)	I use it but don't particularly like it	I don't like it	I'm unfamiliar with it
Absolut	67%	6%	42%	19%	33%	5%	14%	13%
Brand Category Norm (268 Ads)	51%	7%	27%	17%	49%	2%	7%	39%
Brand Index	131	86	156	113	67	250	200	33

The brand disposition information is asked at the beginning of the MRI Starch survey, before respondents are asked about specific ads. The brand disposition questions do not reference an ad but rather ask the respondent about the brand itself.

Actions Taken and Brand category norms are created using a fixed average of available data from 10/08 to 12/09.

[=] Ad has fewer than 4 words [-] Ad has fewer than 50 words | Adnorms Grp = Online10CT (10/2008-12/2009) | Run Date: Aug 24, 2010 2:41 PM

Starch claims that (1) the pulling power of various aspects of the ad can be assessed through the control offered, (2) the effectiveness of competitors' ads can be compared through the norms provided, (3) alternative ad executions can be tested, and (4) readership scores are a useful indication of consumers' *involvement* in the ad or campaign. (The theory is that a reader must read and become involved in the ad before the ad can communicate. To the degree that this readership can be shown, it is a direct indication of effectiveness.)

Of these claims, perhaps the most valid is the ability to judge specific aspects of the ad. Some researchers have criticized other aspects of the Starch recognition method (as well as other recognition measures) on the basis of problems of false claiming, interviewer sensitivities, and unreliable scores:

1. *False claiming.* Research shows that in recognition tests, respondents may claim to have seen an ad when they did not. False claims may be a result of having seen similar ads elsewhere, expecting that such an ad

would appear in the medium, or wanting to please the questioner. Interest in the product category also increases reporting of ad readership. Whether this false claiming is deliberate or not, it leads to an overreporting of effectiveness. On the flip side, factors such as interview fatigue may lead to an underreporting bias—that is, respondents not reporting an ad they did see.

2. *Interviewer sensitivities.* Anytime research involves interviewers, there is a potential for bias. Respondents may want to impress the interviewer or fear looking unknowledgeable if they continually claim not to recognize an ad. There may also be variances associated with interviewer instructions, recordings, and so on, regardless of the amount of training and sophistication involved.
3. *Reliability of recognition scores.* Starch admits that the reliability and validity of its readership scores increase with the number of insertions tested, which essentially means that to test just one ad on a single exposure may not produce valid or reliable results.

page 618

In sum, despite critics, the Starch readership studies continue to dominate the posttesting of print ads. The value provided by norms and the fact that multiple exposures can improve reliability and validity may underlie the decisions to employ this methodology. In addition, the long-standing reliability and trustworthiness of the organization definitely comes into play, as the Starch methodology has been used for over 85 years, and has continued to evolve with the times.

Recall Tests There are several tests to measure recall of print ads. Perhaps the best known of these are the Ipsos-ASI Next*Correct test and the Gallup & Robinson Magazine Impact Research Service (MIRS) described in Figure 18–14. (A number of other companies, including Kantar Millward Brown and Decision Analyst, offer similar services.) These **recall tests** are similar to those discussed in the section on pretesting broadcast ads in that they attempt to measure recall of specific ads.

IGURE 18–14

G&R Magazine Impact Research Service (MIRS)

Objective
Tracking recall of advertising (and client's ads) appearing in magazines to assess performance and effectiveness.
Method
Test magazines are placed in participants' homes and respondents are asked to read the magazine that day. Test ads are embedded with editorial content and other ads. A telephone interview is conducted the next day to assess recall of ads, idea communication, persuasion, brand rating, and ad liking. Both business-to-business and consumer magazine ads can be tested.
Output
Measures provided: • Norms. • Verbatim respondent playback. • Performance summary. • Diagnostics.

In addition to having the same interviewer problems as recognition tests, recall tests have other disadvantages. The reader's degree of involvement with the product and/or the distinctiveness of the appeals and visuals may lead to higher-than-accurate recall scores, although in general the method may lead to lower levels of recall than actually exist (an error the advertiser would be happy with). Critics contend the test is not strong enough to reflect recall accurately, so many ads may score as less effective than they really are, and advertisers may abandon or modify them needlessly.

On the plus side, it is thought that recall tests can assess the ad's impact on memory. Proponents of recall tests say the major concern is not the results themselves but how they are interpreted. Previous studies have shown that the correlation between ad recall and recognition is very high in both newspapers and magazines.

Posttests of Broadcast Commercials A variety of methods exist for posttesting broadcast commercials. The most common provide a combination of day-after recall tests, persuasion measures, and diagnostics. Test marketing and tracking studies, including single-source methods, are also employed.

Day-After Recall Tests The most popular method of posttesting employed in the broadcasting industry for decades was the *Burke Day-After Recall test.* While a number of companies offered day-after recall (DAR) methodologies, the "Burke test" for all intents and purposes became the generic name attached to these tests. While popular, day-after recall tests also had problems, including limited samples, high costs, and security issues (ads shown in test markets could be seen by competitors). Because of their common usage, numerous studies have been conducted to determine the efficacy of DAR tests. Although these studies have been conducted quite some time ago, the conclusions seem to be relevant today and merit consideration here. In addition, the following disadvantages with recall tests were also suggested:

1. DAR tests may favor unemotional appeals because respondents are asked to verbalize the message. Thinking messages may be easier to recall than emotional communications, so recall scores for emotional ads may be lower.[15] A number of other studies have also indicated that emotional ads may be processed differently from thinking ones; some ad agencies, for example, Leo Burnett and BBDO Worldwide among others, have gone so far as to develop their own methods of determining emotional response to ads.[16]
2. Program content may influence recall. The programs in which the ad appears may lead to different recall scores for the same brand. The net result is a potential inaccuracy in the recall score and in the norms used to establish comparisons.
3. A pre-recruited sample may pay increased attention to the program and the ads contained therein because the respondents know they will be tested the next day. This testing effect would lead to a higher level of recall than really exists.

4. In addition, studies have shown that recall is a measure that the ad has been received, but not necessarily accepted, and not predictive of sales.[17,18]

The major advantage of day-after recall tests is that they are field tests. The natural setting is supposed to provide a more realistic response profile. These tests are also popular because they provide norms that give advertisers a standard for comparing how well their ads are performing. In addition to recall, a number of different measures of the commercial’s effectiveness are now offered, including persuasive measures and diagnostics. (The Burke test itself no longer exists under that name.)

Persuasive Measures As noted earlier in our discussion of pretesting broadcast commercials, a number of research firms now offer measures of a commercial’s persuasive effectiveness. Some of the services offer additional persuasion measures, including purchase, intent, and frequency-of-purchase criteria.

Diagnostics In addition to measuring recall and persuasion, copy testing firms also provide diagnostic measures. These measures are designed to garner viewers’ evaluations of the ads, as well as how clearly the creative idea is understood and how well the proposition is communicated. Rational and emotional reactions to the ads are also examined. A number of companies offer diagnostic measures, including G&R and Kantar Millward Brown, among many others.

Comprehensive Measures While each of the measures just described provides specific input into the effectiveness of a commercial, many advertisers are interested in more than just one specific input. Thus, some companies provide comprehensive approaches in which each of the three measures just described can be obtained through one testing program.

Test Marketing Many companies conduct tests designed to measure their advertising effects in specific test markets before releasing them nationally. The markets chosen are representative of the target market. For example, a company may test its ads in Portland, Oregon; San Antonio, Texas; or Buffalo, New York, if the demographic and socioeconomic profiles of these cities match the product's target market. A variety of factors may be tested, including reactions to the ads (for example, alternative copy points), the effects of various budget sizes, or special offers. The ads run in finished form in the media where they might normally appear, and effectiveness is measured after the ads run.

The advantage of test marketing of ads is realism. Regular viewing environments are used and the testing effects are minimized. A high degree of control can be attained if the test is designed successfully. For example, an extensive test market study was designed and conducted by Seagram and Time, Inc. over a period of three years to measure the effects of advertising frequency on consumers' buying habits. This study demonstrated just how much could be learned from research conducted in a field setting but with some experimental controls. It also showed that proper research can provide strong insights into the impact of ad campaigns. (Many advertising researchers consider this study one of the most conclusive ever conducted in the attempt to demonstrate the effects of advertising on sales.)

The Seagram study also reveals some of the disadvantages associated with test market measures, not the least of which are cost and time. Few firms have the luxury to spend three years and hundreds of thousands of dollars on such a test. In addition, there is always the fear that competitors may discover and intervene in the research process.

A number of companies, including Procter & Gamble and Toyota, have test-marketed interactive commercials. Reckitt—the world's largest manufacturer of household cleaning products—and Whirlpool have joined efforts to test iTV ads. Customers were offered three different enticements to interact with the campaign: (1) register to win a Whirlpool dishwasher, (2) register for free samples of Finish Dishwater Freshener, or (3) order money-off coupons for Finish Dishwater Tablets. After eight months of testing, Reckitt reported that the target goal of 35,000 responses was exceeded.[19]

Test marketing can provide substantial insight into the effectiveness of advertising if care is taken to minimize the negative aspects of such tests.

Single-Source Tracking Studies Since the 1980s the focus of many research efforts has been on single-source tracking methods. **Single-source tracking methods** track the behaviors of consumers from the television set to the supermarket checkout counter. Participants in a designated area who have cable TV and agree to participate in the studies use optical scanning equipment that identifies their household and gives the research company their demographics. The households are split into matched groups; one group receives an ad while the other does not, or alternate ads are sent to each. Their purchases are recorded from the bar codes of the products bought. Commercial exposures are then correlated with purchase behaviors.

Earlier we mentioned the use of single-source ad research in pretesting commercials. The single-source method can also be used effectively to posttest ads, allowing for a variety of dependent measures and tracking the effects of increased ad budgets and different versions of ad copy—and even ad effects on sales.

A 10-year study conducted by Information Resources's BehaviorScan service demonstrated long-term effects of advertising on sales. The study examined copy, media schedules, ad budgets, and the impact of trade promotions on sales in 10 markets throughout the United States and concluded that advertising can produce sales growth as long as two years after a campaign ends.[20] (The study also concluded that results of copy recall and persuasion tests were unlikely to predict sales reliably.) A number of single-source methods have been used, including the Nielsen Company's Catalina (Exhibit 18–8).

XHIBIT 18–8

Neilson Catalina offers clients the opportunity to use single- source measurement.

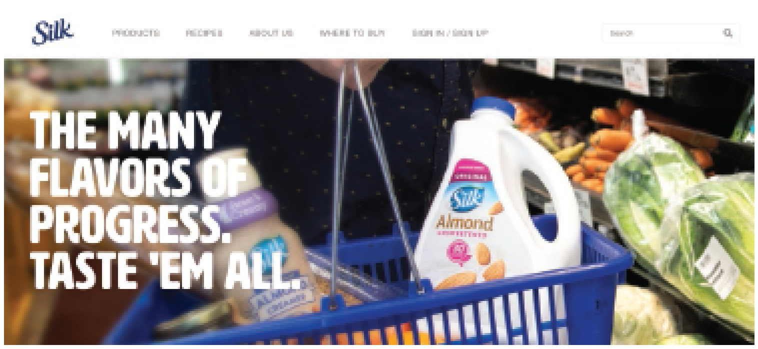
ource: Nielsen Catalina Solutions

Many advertisers believe these single-source measures will change the way research is conducted due to the advantages of control and the ability to measure directly the ads' effects on sales. A number of major corporations and ad agencies are now employing this method, including companies and their agencies in the automotive, entertainment, financial services, packaged goods, and pharmaceutical industries among others. Nielsen has used Catalina to improve the marketing efforts of Chobani Yogurt, Yahoo, and Silk Brands, among others (Exhibit 18–9).

XHIBIT 18–9

Silk Brands is one of many who employ single source measurement.

ource: Groupe Danone



Whereas single-source testing is a valuable tool, it still has some problems. One researcher says, "Scanner data focus on short-term sales effects, and as a result capture only 10 to 30 percent of what page 621

advertising does."[21] Others complain that the data are too complicated to deal with, as an overabundance of information is available. Still another disadvantage is the high cost of collecting single-source data. While the complexity of single-source data resulted in a slow adoption rate, this method of tracking advertising effectiveness became widely adopted in the 1990s by the research companies mentioned earlier (G&R, Kantar Millward Brown, and Ipsos-ASI).

Tracking Print/Broadcast Ads One of the more useful and adaptable forms of posttesting involves tracking the effects of the ad campaign by taking measurements at regular intervals. **Tracking studies** have been used to measure the effects of advertising on awareness, recall, interest, and attitudes toward the ad and/or brand as well as purchase intentions. (Ad tracking may be applied to print and digital ads.) Personal interviews, phone surveys, mall intercepts, and online surveys have been used. Tracking studies yield perhaps the most valuable information available to the marketing manager for assessing current programs and planning for the future.

The major advantage of tracking studies is that they can be tailored to each specific campaign and/or situation. A standard set of questions can track effects of the campaign over time or through the consumer purchase funnel. In a study by the research organization Yankelovich and the Television Bureau of Advertising (TVB), it was shown that the effectiveness of advertising depends on the product category and where the consumer is in the purchase funnel. The advertising medium impact varied as the consumer moved through the stages.[22] The effects of various media can also be determined, although with much less effectiveness. Tracking studies have also been used to measure the differential impact of different budget sizes, the effects of flighting, brand or corporate image, and recall of specific copy points. As you will see later in the chapter, however, it is often difficult to quantify some of the measures suggested. Finally, when designed properly, as shown in Figure 18–15, tracking studies offer a high degree of reliability and validity.

IGURE 18–15

-actors That Make or Break Tracking Studies

1. Properly defined objectives.
2. Alignment with sales objectives.
3. Properly designed measures (e.g., adequate sample size, maximum control over interviewing process, adequate time between tracking periods).
4. Consistency through replication of the sampling plan.
5. Random samples.
6. Continuous interviewing (that is, not seasonal).
7. Evaluation of measures related to behavior (attitudes meet this criterion; recall of ads does not).
8. Critical evaluative questions asked early to eliminate bias.
9. Measurement of competitors' performance.
10. Skepticism about questions that ask where the advertising was seen or heard (TV always wins).
11. Building of news value into the study.
12. "Moving averages" used to spot long-term trends and avoid seasonality.
13. Data reported in terms of relationships rather than as isolated facts.
14. Integration of key marketplace events with tracking results (e.g., advertising expenditures of self and competitors, promotional activities associated with price changes in ad campaigns, introductions of new brands, government announcements, changes in economic conditions).

Some of the problems of recall and recognition measures are inherent in tracking studies, since many other factors may affect both brand and advertising recall. Despite these limitations, however, tracking studies are a very effective way to assess the effects of advertising campaigns.

page 622

In summary, you can see that each of the testing methods considered in this chapter has its strengths and its limitations. You may wonder: Can we actually test advertising effectiveness? What can be done to ensure a valid, reliable test? The next section of this chapter suggests some answers.

ESTABLISHING A PROGRAM FOR MEASURING ADVERTISING EFFECTS

There is no surefire way to test advertising effectiveness. However, in response to pressures to determine the contribution of ads—traditional and online—to the overall marketing effort, steps are being taken to improve this measurement task. Let's begin by reviewing the major problems with some existing methods and then examine possible improvements.

Problems with Current Research Methods

LO 18-3

When current testing methods are compared to the criteria established by PACT (see Figure 18–3) and Making Measurement Make Sense (Figure 18–4), it is clear that some of the principles important to good testing can be accomplished readily, whereas others require substantially more effort. For example, PACT principle 6 (providing equivalent test ads) should require a minimum of effort. The researcher can easily control the state of completion of the test communications. Also fairly easy are principles 1 and 2 (providing measurements relative to the objectives sought and determining *a priori* how the results will be used). In regard to Figure 18–4, and digital measurement, some of the principles (for example, principle 2, the currency issue, and 5, comparable measures) may not be difficult to achieve.

We have seen throughout this text that each promotional medium, the message, and the budget all consider the marketing and communications objectives sought. The integrated marketing communications planning model establishes the roles of these elements. So by the time one gets to the measurement phase, the criteria by which these programs will be evaluated should simply fall into place.

Slightly more difficult are PACT principles 3, 5, and 8, although again these factors are largely in the control of the researcher. Principle 3 (providing multiple measurements) may require little more than budgeting to make sure more than one test is conducted. At the most, it may require considering two similar measures to ensure reliability. Likewise, principle 5 (exposing the test ad more than once) can be accomplished with a proper research design. Finally, principle 8 (sample definition) requires little more than sound research methodology; any test should use the target audience to assess an ad's effectiveness. You would not use a sample of nondrinkers to evaluate new liquor commercials. Likewise principles 3 and 4 on the digital side might take some time and give and take, but can be accomplished.

The most difficult factors to control—and the principles that may best differentiate between good and bad testing procedures—are PACT requirements 4, 7, and 9. Fortunately, however, addressing each of these contributes to the attainment of the others. On the digital side, developing "viewable impressions" is proving to be a difficult task.

While it is important that marketers' attempts to measure effectiveness be guided by all of the principles stated, the research should be guided by a model of human response to communications that encompasses reception, comprehension, and behavioral response. It is the best starting point, in our opinion, because it is the principle least addressed by practicing researchers. If you recall, Chapter 5 proposed a number of models that could fulfill this requirement. Yet even though the models have existed for quite some time, few if any common research methods attempt to integrate them into their methodologies. Most current methods do little more than provide recall scores, despite the fact that many researchers have shown recall to be a poor measure of effectiveness. Models that do claim to measure such factors as attitude change or brand preference change are often fraught with problems that severely limit their reliability. Once again, a problem with digital metrics is that they often don't provide insights into the consumer response processes because the measures are behavioral and provide no insights into the emotional side. An effective measure must include some relationship to the communication process.

It might seem at first glance that PACT principle 7 (providing a nonbiasing exposure) would be easy to accomplish. But lab measures, while offering control, are artificial and vulnerable to testing effects. And field measures, although more realistic, often lose control. The Seagram and Time study may have the best of both worlds, but it is too large a task for most firms to undertake. Some of the improvements associated with the single-source systems help solve this problem. In addition, properly designed ad tracking studies provide truer measures of the impact of the communication. As technology develops and more attention is paid to this principle, we expect to see improvements in methodologies soon.

Last but not least is PACT principle 9, the concern for reliability and validity. This principle must carry over to online measures as well. Most of the measures discussed are lacking in at least one of these criteria, yet these are two of the most critical distinctions between good and bad research. If a study is properly designed, and by that we mean it addresses principles 1 through 8, it should be both reliable and valid. Studies designed to measure the impact of online advertising also need to adhere to the PACT principles as well as the Making Measurement Make Sense ones. As you have just read, many of the larger research firms mentioned here are now applying platforms to measure both traditional and digital ads.

Essentials of Effective Testing

Simply put, good tests of advertising effectiveness must address the nine principles established by PACT. One of the easiest ways to accomplish this is by following the decision sequence model in formulating promotional plans.

- *Establish communications objectives.* We have stated that except for a few instances (most specifically direct-response advertising), it is nearly impossible to show the direct impact of advertising on sales. So the marketing objectives established for the promotional program are not usually good measures of communication effectiveness. For example, it is very difficult (or too expensive) to demonstrate the effect of an ad on brand share or on sales. On the other hand, attainment of communications objectives can be measured and leads to the accomplishment of marketing objectives.
- *Use a consumer response model.* Early in this text we reviewed the hierarchy of effects models and cognitive response models, which provide an understanding of the effects of communications and lend themselves to achieving communications goals. Many companies including Honda, General Motors, Sprint, and others use hierarchy models to establish objectives and assess effectiveness in both the traditional and online environments.

- *Use both pretests and posttests.* From a cost standpoint—both actual cost outlays and opportunity costs—pretesting makes sense. It may mean the difference between success or failure of the campaign or the product. But it should work in conjunction with posttests, which avoid the limitations of pretests, use much larger samples, and take place in more natural settings. Posttesting may be required to determine the true effectiveness of the ad or campaign.
- *Use multiple measures.* Many attempts to measure the effectiveness of advertising focus on one major dependent variable—perhaps sales, recall, or recognition. As noted earlier in this chapter, advertising may have a variety of effects on the consumer, some of which can be measured through traditional methods and others that require page 624 updated thinking (recall the discussion on physiological responses). For a true assessment of advertising effectiveness, a number of measures may be required. The Ogilvy Award winners mentioned earlier all employed multiple measures to track the effects on communications objectives.
- *Understand and implement proper research.* It is critical to understand research methodology. What constitutes a good design? Is it valid and reliable? Does it measure what we need it to? There is no shortcut to this criterion, and there is no way to avoid it if you truly want to measure the effects of advertising.

MEASURING THE EFFECTIVENESS OF OTHER PROGRAM ELEMENTS

Throughout this text, we have discussed how and when promotional program elements should be used, the advantages and disadvantages of each, and so on. In many chapters we have discussed measures of effectiveness used to evaluate these programs. In the final section of this chapter, we add a few measures that were not discussed earlier.

Measuring the Effectiveness of Sales Promotions

LO 18-4

Sales promotions are not limited to retailers and resellers of products. Sports marketers have found them a very effective way to attract crowds and have been able to measure their relative effectiveness by the number of fans attending games. Major League Baseball teams have seen their attendance increase for those games in which promotions are offered.

A number of organizations measure the impact of sales promotions. The National Shopper Lab (NSL) tracks the effectiveness of promotions leveraging data collected from more than 80 million consumers. NSL captures the data from shoppers' weekly transactions through Shopper ID, UPC, store, date, price, and coupon. NSL is a division of GfK—an international research company discussed at various points throughout this text. Promotion Decisions Inc. examines the impact of freestanding inserts (FSIs) in Figure 18–16.

IGURE 18–16

Measuring the Effects of FSIs

A study by Promotion Decisions Inc. examined the actual purchase data of users and nonusers of 27 coupon promotions in its National Shopper Lab (75,000 households) over a period of 18 months. The findings are as follows:

- FSI coupons generated significant trial by new and lapsed users of a product (53 percent).
- Repeat purchase rates were 11.8 percent higher among coupon redeemers than nonredeemers.
- 64.2 percent of repeat volume among coupon redeemers was without a coupon.

- There was no significant difference in share of volume between buyers who used coupons and those who did not.
- Coupons returned between 71 and 79 percent of their cost within 12 weeks.
- Full-page ads provided higher redemption rates, incremental volume, redemption by new users, and a higher number of repeat buyers than half-page ads.
- Consumers who used coupons were brand loyal.

Other measures of sales promotions are also available. For example, a number of research companies provide retailers visibility into the factors that influence the effectiveness of all of their promotions, allowing the retailer to make the right decision about the right products to promote using promotional vehicles and understanding the required inventory levels. Other companies have used awareness tracking studies and count the number of inquiries, coupon redemptions, and sweepstakes entries. They also track sales during promotional and nonpromotional periods while holding other factors constant. The Nielsen Company's ScanTrack methodology helps marketers answer the questions regarding how many promotions to do and how this promotion compares to competitors', and what overall sales lifts are associated with individual promotions using scanner data. An article by SAS concludes that retailers that make promotion decisions without considering measurable customer behavior are at a great disadvantage in the marketplace.[23]

One interesting and useful technology designed to track the effectiveness of sales promotions at the point of sale is offered by Shopper Trak. Shopper Trak employs heat map technology in the store that views traffic; whether a person is coming or going; calculates the shopper's height (to differentiate between adults and children); and gauges traffic patterns and sales conversions, among other metrics. The system helps retailers evaluate the effectiveness of promotions or displays located throughout the store.[24]

Measuring the Effectiveness of Nontraditional Media

In Chapter 13, we noted that one of the disadvantages of employing nontraditional media is that it is usually difficult to measure the effectiveness of the programs. But some progress has been made, as shown in these examples:

- *The effectiveness of print versus online flyers*—In an article published in the *Journal of Advertising Research,* Marco Ieva, Juan Carlos Gazquez-Abad, Ida D'ttoma, and Christina Ziliani examined the difference between online and traditional flyers' effects on memory and purchase behaviors. The measures were tracked in 37 supermarket chains over 9,902 consumers. The results indicated that there was no difference between the print and online flyers' impact on consumers.[25]
- *The effectiveness of parking lot–based media.* In Chapter 13, we discussed advertising on parking lot signage and other areas to attempt to reach selective demographic groups. Now the Traffic Audit Bureau (TAB) is tracking the effectiveness of this form of advertising to give advertisers more reliable criteria on which to base purchase decisions. The TAB data verify ad placements, while the media vendors have employed Experian Simmons Market Research and Nielsen Media Research to collect ad impressions and advertising recall information. These measures are combined with sales tracking data to evaluate the medium's effectiveness.
- *The effects of in-store radio and television.* Interactive Market Systems (IMS) introduced software that enables clients to measure the effectiveness of in-store radio. The company planned to introduce similar software designed to measure in-store television advertising effectiveness.[26]
- *The effectiveness of other media.* A number of companies provide effectiveness measures to determine the impact of package designs, POP displays, trade show exhibits, and the like. Nielsen Entertainment and Massivemedia now offer a service to measure video game advertising effectiveness as well as that of other outdoor media. While it is not

possible to list them all here, suffice it to say that if one wants to measure the impact of various IMC elements, the resources are available.

Measuring the Effectiveness of Sponsorships

In earlier chapters we discussed the growth in sponsorships and the reasons why organizations have increased their investments in this area. Along with the increased expenditures have come a number of methods for measuring the impact of sponsorships. Essentially, measures of sponsorship effectiveness can be categorized as exposure-based methods or tracking measures:[27]

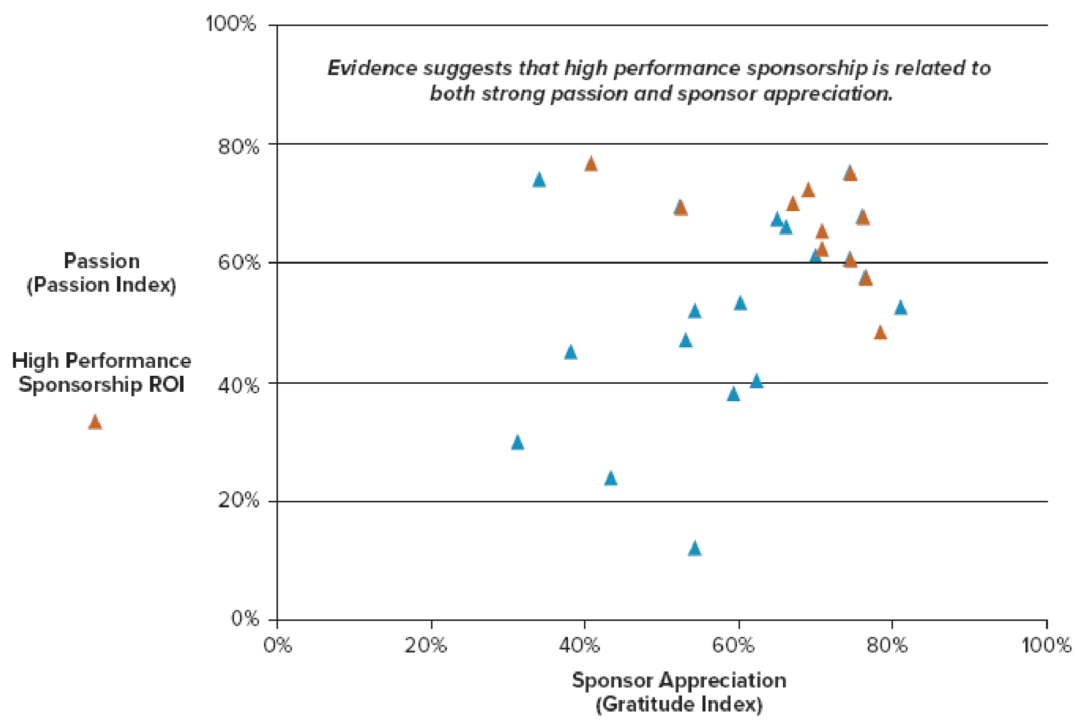
- *Exposure methods.* Exposure methods can be classified as those that monitor the quantity and nature of the media coverage obtained for the sponsored event and those that estimate direct and indirect audiences. While commonly employed by corporations, scholars have page 626 heavily criticized these measures. For example, Michel Pham argues that media coverage is not the objective of sponsorships and should not be considered as a measure of effectiveness. He argues that the measures provide no indication of perceptions, attitude change, or behavioral change and should therefore not be considered as measures of effectiveness.[28]
- *Tracking measures.* These measures are designed to evaluate the awareness, familiarity, and preferences engendered by sponsorship based on surveys. A number of empirical studies have measured recall of sponsors' ads, awareness of and attitudes toward the sponsors and their products, and image effect, including brand and corporate images.

A number of companies now measure the effectiveness of sports and other sponsorships. For example, companies assign a value referred to as media equivalency and assign a monetary value to the amount of exposure the sponsor receives during the event. They review broadcasts and add up the number of seconds a sponsor's product name or logo can be seen clearly (e.g., on signs or shirts). A total of 30 seconds is considered the equivalent of a 30-second commercial. (Such measures are of questionable validity.)

As with all other IMC touch points, marketers would like to measure the ROI of event sponsorships. But—like these other means of communication—this is usually not possible, so the focus will be on other more communication-oriented objectives. Figure 18–17 shows a Sponsorship Performance Matrix used by sponsor effectiveness measurement company SponsorMap. The vertical axis of this matrix measures the passion the target audience feels toward the sponsoring organization. The more emotionally attached to the property (sporting team, event, etc.), the higher the Passion Index. The horizontal axis measures sponsor appreciation—the Gratitude Index. According to SponsorMap, companies can achieve more appreciation by being more active, thus creating more goodwill. As can be seen in Figure 18–7, the sponsor would want to be in the upper right-hand corner to achieve maximum effectiveness.

IGURE 18–17

Sponsorship Performance Matrix



Performance research measures impact on brand awareness and image shifts. Tracking the number of website visits and buzz generated are also criteria that have been used.

While each of these measures has its advantages and disadvantages, most do not go far enough. As noted by John Nardone and Ed See, most marketers limit their sponsorship evaluations to brand awareness and impressions. The key question that needs to be asked, they say, is "how do you do sponsorships that build brand equity and maintain financial responsibility?"[29] We suggest using several in assessing the impact of sponsorships. In addition to those mentioned here, the eight-step process suggested in Figure 18–18 could be used to guide these evaluations.

page 627

IGURE 18–18

Eight Steps to Measuring Event Sponsorship

1. Narrowly define objectives with specifics.
2. Establish solid strategies against which programming will be benchmarked and measure your programming and effectiveness against the benchmark.
3. Set measurable and realistic goals; make sure everything you do supports them.
4. Enhance, rather than just change, other marketing variables.
5. Don't pull Marketing Plan 101 off the shelf. Programming should be crafted to reflect the particulars of your company's constituencies and target audiences.
6. Define the scope of your involvement. Will it involve multiple areas within the company? Who internally and externally constitutes the team?
7. Think "long term." It takes time to build brand equity. Also, think of leveraging your sponsorship through programming for as long as possible, before and after the event.
8. Build evaluation and a related budget into your overall sponsoring program. Include items such as pre- and postevent attitude surveys, media analyses, and sales results.

Measuring the Effectiveness of Other IMC Program Elements

Many of the organizations mentioned in this chapter offer research services to measure the effectiveness of specific promotional program elements. As we noted at the outset of this chapter, the increased use of integrated marketing communications programs has led to more interest in determining the synergistic effects of all program elements. A review of the Ogilvy Award winners from 1993 to date demonstrates the increased integration of additional media (as opposed to specifically the best advertising campaign) and the value of measuring their contribution to the program's success. Also departing from the specific focus on advertising are the awards given by the London-based Institute of Practitioners, which has opened the competition to nontraditional media as well as public relations, sales promotions, and other entries.

As noted, a number of studies have been implemented to determine the combined effects of two or more media as well as their synergistic impact. The number of studies being designed to specifically measure synergistic effects continues to increase—most of which demonstrate a higher effectiveness when multiple media are employed.

As previously stated, a number of research firms such as Kantar Millward Brown and Nielsen Media Research are applying traditional advertising effectiveness measures to their online advertising to assess their overall communications effects.[30]

Perhaps the major challenge facing the adoption of an IMC approach is determining the effectiveness and contribution of various elements of the IMC program. A number of academic studies have offered insights into this endeavor, as has at least one company. For example, Dinner et al. examined the "presence, magnitude, and carryover of cross-channel effects for online advertising (display and search) and traditional media."[31] Their conclusion was that cross-channel effects do exist and are important.

One very effective approach to measuring the impact of the IMC program is that provided by the marketing communications research company Integration. The company's approach is based on the belief that integrated marketing communications improve both the efficiency and the effectiveness of a campaign. Integration contends that most traditional measurement techniques focus only on the former of these (see Figure 18–19). Noting the increased demand for marketing managers to prioritize the media vehicles used to promote their brands, Integration developed Market Contact Audit (MCA) to measure both the efficiency and the effectiveness of media used to establish contacts with consumers. By measuring consumers' understanding, evaluation, and perceptions of the contacts and their association with the brand, MCA allows marketers to assess the overall effectiveness as well as the relative contribution of individual IMC elements.[32]

IGURE 18–19

Measuring Effectiveness, Not Only Efficiency

The most important issue for IMC planning is identifying the most appropriate contact mix, that is, which contacts to prioritize.

Questions: Are we . . .	
Being effective:	1. Doing the right things?
Being efficient:	2. Doing things right?
Current techniques measure only efficiency	
Reach, frequency, GRPs (media surveys)	
Weight of market activity (stochastic)	
MCA by integration measures effectiveness and efficiency	
Influence of contacts	
Consumer brand experience	

Despite these efforts, it appears we still have a long way to go to determine the overall effectiveness of an IMC program. All the advertising effectiveness measures discussed here have their inherent strengths and weaknesses. They offer the advertiser some information that may be useful in evaluating the effectiveness of promotional efforts. While not all promotional efforts can be evaluated effectively, progress is being made.

Summary

This chapter introduced you to issues involved in measuring the effects of advertising and promotions. These issues include reasons for testing, reasons companies do not test, and the review and evaluation of various research methodologies. We arrived at a number of conclusions: (1) Advertising research to measure effectiveness is important to the promotional program, (2) not enough companies test their ads, and (3) problems exist with current research methodologies. In addition, we reviewed the criteria for sound research and suggested some ways to accomplish effective studies.

All marketing managers should want to know how well their promotional programs are working. This information is critical to planning for the next period, since program adjustments and/or maintenance are based on evaluation of current strategies. Problems often result when the measures taken to determine such effects are inaccurate or improperly used.

This chapter demonstrated that testing must meet a number of criteria (defined by PACT and Making Measurement Make Sense) to be successful. These evaluations should occur both before and after the campaigns are implemented.

A variety of research methods were discussed, many provided by syndicated research firms such as Ipsos-ASI, MSW•ARS, Kantar Millward Brown, and the Nielsen Co. All of these companies have expanded their service offerings to assist in the evaluation of online and other advertising platforms. Many companies have developed their own testing systems. There has been an increase in testing through the Internet.

Single-source research data were discussed. These single-source systems offer strong potential for improving the effectiveness of ad measures in the future, since commercial exposures and reactions may be correlated to actual purchase behaviors. However, to date their use has not met expectations.

It is important to recognize that different measures of effectiveness may lead to different results. Depending on the criteria used, one measure may show that an ad or promotion is effective while another states that it is not. This is why clearly defined objectives and the use of multiple measures are critical to determining the true effects of an IMC program.

page 629

Key Terms

vehicle option source effect 601
pretests 601
posttests 601
laboratory tests 602
testing bias 602
field tests 603
Positioning Advertising Copy Testing (PACT) 603
concept testing 604
comprehension and reaction tests 606
consumer juries 606
halo effect 607
A/B testing 607
portfolio tests 608
mock magazine test 608
Flesch formula 609
theater testing 610
on-air test 611
recall 611
pupillometrics 611
electrodermal response (EDR) 611
eye tracking 612
electroencephalographic (EEG) measures 614
alpha activity 614
hemispheric lateralization 614
inquiry tests 615
split-run tests 615
recognition method 616
recall tests 618
single-source tracking methods 620
tracking studies 621

Discussion Questions

1. In the decision sequence model presented in Chapter 1, the final stage is measuring effectiveness of the IMC program to provide feedback. This feedback is critical to the future planning process. Yet research shows that many companies do not conduct effectiveness research. Explain why they do not and why it is important that they do. (LO 18-1)

2. Digital and Social Media Perspective 18-1 discusses the fact that some marketers do not trust the effectiveness measures that are available to them, and thus do not attempt to measure the impact of their programs. Are they right, or wrong? Provide examples to support your position. (LO 18-3)

3. Creatives have long argued that measuring the effectiveness of their ads would stifle the creative process and lead to less creativity and more objective seeking ads. The outstanding creative ads of the past would no longer be part of the future, they say. Some recent articles have contended that the ads of today are not creative and tend to copy each other. Take a side on this position, and cite examples to support it. (LO 18-1)

4. Digital and Social Media Perspective 18-2 discusses the agreement by practitioners and academics that physiological measures of effectiveness are valid and a value to advertisers. Discuss some of these physiological measures and the value they bring to marketers attempting to measure effectiveness. Cite examples. (LO 18-2)

5. What is single source measurement? Why does is appear to offer marketers an advantage over other measures of effectiveness? Are there any disadvantages to using single source measures? (LO 18-4)

6. A/B testing has been used by marketers for decades to evaluate two different versions of finished ads and/or web pages. Some experts now believe that A/B testing would be more effective if employed earlier in the campaign development process. Explain what they mean by this, and argue for or against this position. (LO 18-1)

7. Explain why it is so difficult to measure the overall IMC program effectiveness. (LO 18-3)

8. This chapter discussed the Ogilvy Awards. Describe what these awards are and how they have changed over time. Give examples of companies that have won these awards (you can find past winners online) and why they have won them. (LO 18-1)

9. Explain why it is important to pretest. When should this testing take place? Give examples of what might make ads or commercials that are pretested more successful. (LO 18-1)

10. Digital advertisers seem to want to have their own effectiveness measures, while eschewing those of traditional advertisers. What are the pros and cons of this position? Argue for one side or the other. (LO 18-4)

19 International Advertising and Promotion

Source: DisobeyArt/Shutterstock

Learning Objectives

LO 19-1 | Describe the role and importance of international marketing and promotion.