



# *Short History* of Television Copytesting

by Charles E. Young

© 2001 Ameritest/CY Research, Inc.



4841 Tramway Ridge NE, Albuquerque, NM 87111, [www.ameritest.net](http://www.ameritest.net), [info@ameritest.net](mailto:info@ameritest.net)

# A Short History of Television Copytesting

Charles Young  
Ameritest/CY Research

In the late 1950's and early 1960's, television was just over a decade into its commercialization stage, about as far as the internet is today. In those days there were just three networks. Programming was in black and white. And a modern television viewer just might find the content tedious in terms of its visual pacing and heavy on dialogue compared to the sophisticated cinematography of today.

Television advertising was different then, too. The basic unit of advertising was longer — the 60-second commercial. Many brands advertised in sole sponsor shows without commercial clutter. And there were fewer brands doing television advertising — but with more commercials surrounding the brand with a variety of messages.

Let me illustrate that last point. Some years ago, one of the advertising agencies I worked for undertook a bit of advertising archaeology, reviewing the ancient history of two Proctor and Gamble (P & G) brands which they had handled from their very beginning. We were quite surprised to learn how much things had changed since the golden age of television. In 1959, P & G launched *Mr. Clean* household cleanser nationally and, during the first eighteen months of its introduction, aired 35 different sixty-second executions. In 1961, when P & G launched *Head & Shoulders* shampoo, 27 distinct executions ran during the first eighteen months. Today, either brand would be lucky to produce two to three commercials in a given year and one would probably be a :15.

The reason for this profligate rate of commercial production was quite simple. According to historical records which we tracked down in the 4A's library in New York, in 1960 an average network-quality sixty-second commercial would cost you \$10,000 to produce — one-fortieth the cost of today's production. An "A" list director would cost you around \$3,500 — a tiny fraction of today's fees. Not surprisingly, given the low cost of commercial production, doing research on creative effectiveness was considered a relatively simple and straightforward filtering process.

In the years since, however, various methods of research have been developed to manage the risk associated with increasingly expensive media budgets. This category of research has historically been known as "copytesting" or, more correctly, "pre-testing." Implicitly, all of these methods are intended to be *predictive* of commercial performance in some way. The other major form of quantitative advertising research which also developed over this time period involves tracking the effects of television advertising once it has actually aired and separating those effects from other variables in the marketing mix, as exemplified by Millward Brown's continuous tracking service offering (in Brown, 1984, 1987). This form of research is *explanatory* in nature, rather than predictive, and provides

valid surrogate measures for the return on the advertising investment. It's an equally fascinating story, but beyond the scope of this paper.

Much of the history of copytesting resembles the fable of the three blind men describing the elephant with recall or persuasion replacing snake or tree trunk as the competing descriptions of the advertising animal. So, not surprisingly, clients have been much confused by the various pictures painted by the different copytesting systems to describe how advertising works. Yet most of the widely-used approaches are probably valid up to a point. In the past twenty years, suppliers have produced an endless series of validation charts and regression lines with high r-squared statistics over the years in support of their claim to having the best copytesting system. It should be noted that this wasn't always the case. As documented in Ostlund, Clancy, and Sapra (1980), industry frustration with the lack of "research hygiene" among copytesting firms had reached a peak level shortly before the advent of the retail scanner revolution.

So why do client and agency frustrations persist, despite the ongoing evolution and "scientification" of copytesting? This frustration comes from overpromise and oversimplification of a complex subject — the question of how an endless variety of television commercials penetrate the human mind to motivate everyday behavior. Really, when you think about it, rocket science looks easy by comparison.

Because the U.S. was the dominant market for television advertising during this period, the story of how television copytesting evolved in the U.S. essentially highlights the various management debates over this time about how advertising is supposed to work — an issue that still holds our attention today with the emergence of global advertising campaigns and the proliferation of media choices. The current time, therefore, is a good vantage point for reviewing where we have been in our thinking. What follows is an attempt to provide a broad overview of the subject so that researchers currently in the field can move forward.

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

## **The Report Card Measures—Filtering the Creative**

Regardless of the issues of inspiration, risk-taking and creative freedom that are involved in the conceptualization of advertising executions, from a management perspective the creative development process is an expensive business process which outputs “products” — i.e., commercials — of highly variable quality. Like any industrial process, control is a function of our ability to measure it. For large firms, those for which advertising is mission-critical to their business, such as P & G, significant quantities of advertising executions are produced every year. Therefore, very simple metrics are needed to provide senior managers with a clear picture of how well the process is working and to provide a check on the quality of the decisions being made by the more junior brand managers and agency teams who are charged with the day-to-day business of making advertising. The key problem, of course, is one of validity — the relationship to sales.

The logic behind the first report card measure for testing television ads, the Day After Recall (DAR) score, is quite simple. For advertising to be effective, it must surely leave some trace behind in the memory of the consumer. This memory effect metric was particularly credible given the traditional argument that advertising is superior to short-term promotions such as couponing only because of its long-term effect on sales. Recall testing, therefore, was interpreted to measure an ad’s ability on air to “break through” into the mind of the consumer to register a message from the brand in long-term memory.

According to Honomichl (1986), DAR testing was first applied to the advertising measurement problem by George Gallup Sr., who built on R & D work done by a Commander Thompson who used it in the training of Navy pilots during WWII. The Compton advertising agency then took up the development of the measure, evaluating an arbitrary range of forgetting periods such as 12, 24, 48 and 72 hours and exploring a range of variables operating in the on-air viewing environment. Compton soon began promoting it to their clients as proof of performance for their work. P & G conducted experiments of its own, became convinced of the usefulness of the measure, and subcontracted the fieldwork to a then small research company in Cincinnati called Burke. With P & G taking the lead, many other advertisers soon followed suit and Burke DAR testing became the dominant copytesting report card measure for the fifties and sixties.

In the seventies, however, some researchers began to question the relationship between recall and sales. According to some reports at the time, P & G reviewed a hundred split-cable test markets that had been conducted over ten years with ads that had been recall-tested and had been unable to find a significant relationship between recall scores and sales response. Not coincidentally, in a major validation study conducted at roughly the same time, Ross (1982) found that persuasion was a better predictor of sales response to advertising than was recall. Much later, Lodish and his colleagues (1995) conducted an even more extensive review of test market results and also failed to find a relationship between recall and sales.

During this period, therefore, the attention of advertising researchers shifted to the problem of measuring advertising persuasiveness. One of the researchers leading the way was Horace Schwerin (in Honomichl, 1986), who pointed out “the obvious truth is that a claim can be well remembered but completely unimportant to the prospective buyer of the product—the solution the marketer offers is addressed to the wrong need.” Schwerin sold his company (and the pre-post shift approach to measuring persuasion that he developed) to ARS, which succeeded in getting it adopted by P & G as the new standard for measuring advertising effectiveness.

Recall continued to be provided as a companion measure, in part to ease the transition with the old-line researchers, with the caveat that recall is important up to some minimal threshold level but persuasion is the more important measure. Again, with the imprimatur of P & G support, the ARS pre-post measure of persuasiveness became the category leader for much of the seventies and eighties (see Blair, 2000 for ARS’s validation work in support of this measure). Alternative post-exposure measures of motivation or persuasiveness were also developed at this time such as the weighted five-point purchase intent scale — the industry gold standard for concept testing (which is currently used by the BASES simulated test market system as well as by several copytesting companies).

Meanwhile, other researchers during the late seventies (summarized in Arnold & Bird, 1982) began to question the validity of recall as a measure of “breakthrough.” To some researchers, the construct of breakthrough is about the ability of the commercial execution to win the fight for attention and get noticed immediately, which is what many Creatives assume is the first task that advertising must perform in creating a sale. Of course, this is not the same as measuring what happens to a commercial after it’s been processed through long-term memory. An important distinction was made between the attention-getting power of the creative execution and how well “branded” the ad was. One of the reasons an ad may fail in a recall test, it was argued, even if it broke through the clutter and garnered a lot of attention, is that the memory of it might simply be filed away improperly in the messy filing cabinets of the mind so that it becomes difficult to retrieve with the standard verbal recall prompts.

This debate ran parallel to the ongoing debate of recall versus recognition as the best approach for tracking advertising awareness in market. These two very different ways of tapping into memory for evidence of advertising awareness can produce very different measurements, usually substantially higher for recognition, as discussed in further detail in Young (2003).

For researchers on the recall side of the debate, the standard approach to measuring ad recall with telephone surveys provides consistency and comparability between the recall results of a pre-test and those of a post-test. Those on the recognition side of the debate remind us that memory is a complex subject, and that more than one memory system of the mind may be involved in determining advertising effectiveness.

Consider the difference between one's failure to recall the name yet ability to recognize the face of someone one has met before. For most people, the fact that one can recognize a face is taken as the more reliable "proof" that one has actually met a person before. In the last few years, the internet has now created a practical research opportunity, unavailable in telephone surveys, to show consumers the "face" of an ad — i.e., a film clip or storyboard — in order to measure ad awareness with recognition-based questions.

Several companies, such as MSW Research, DRI, and Ameritest have developed alternative pre-testing approaches to recall testing as a measure of commercial breakthrough (see Young, 2001). In these systems, attention and branding are measured separately. These approaches involve simulating a cluttered media environment off-air and measuring which ads win the fight for attention without the intervening variable of a day's worth of forgetting time. Recent research, jointly published by Unilever and Ameritest, (in Kastenholz, Young, & Kerr, 2004) has added substance to the debate by showing the findings from an analysis of a large database of commercials that had been tested with both approaches. These results show that recall scores and attention scores are completely uncorrelated, suggesting that these two approaches to measuring "breakthrough" are in fact measuring completely different things. Moreover, recall scores for this set of ads had a strong negative correlation with commercial likeability ratings, confirming what creative directors have been telling researchers about recall tests for many years.

Different approaches were also developed during this period to measure the well-brandedness of a commercial. Ipsos-ASI views brand linkage as the missing variable between recall and recognition and computes a measure of brandedness by looking at the ratio of a commercial's recall score divided by its recognition score (derived by reading a verbal description of an ad to the respondent). The Millward Brown approach to branding uses a five-point rating scale to measure how well a commercial execution is custom tailored to "fit" the brand. A third approach, used by Ameritest, measures branding by tracking a respondent's top-of-mind propensity to use the brand name as the "handle" to retrieve the memory of an advertising execution immediately after exposure in a cluttered media environment. Unfortunately, by using the same branding label to name three different measurement constructs, ad researchers continue to add to the confusion surrounding report card measures. Unilever just reported an analysis of a database of commercials "triple-tested" with all three types of measures, showing that each of the three is measuring something uncorrelated with, and therefore different from, the other two (in Kastenholz, Kerr & Young, 2004).

Finally, a very different approach was used by Millward Brown to determine their report card measures of commercial performance. In the late eighties, working with Unilever, they used a modeling approach to derive an overall “effectiveness index” from various component measures. Instead of starting with an *a priori* theoretical model of how the mind is supposed to interact with advertising, they basically “reverse engineered” the process by running a massive stepwise regression to identify pre-testing measures predictive of advertising tracking results (from their continuous tracking program) and then used a “black box” approach to combine these component measures into an overall predictive score. While not directly validated to sales, this approach had the appeal of creating consistency between the two approaches (pre-testing and in-market tracking) that a company like Unilever uses to grade advertising performance.

Over the last fifty years, category leadership in terms of market share for the commercial firms providing copytesting services has turned on the measure currently in vogue as the “magic number” predicting in-market performance. First, it was Burke’s DAR score. Then it was ARS’s measure of persuasion benchmarked against “fair share.” Currently in the U.S., it is Ipsos-ASI’s “copy effect score,” a composite measure integrating recall and persuasion. And, coming on strong in this highly competitive market, we have Millward Brown, the category leader outside the U.S., with its effectiveness index. The evolution of the entire category can be seen visually in Exhibit 1 as it maps on to the Ameritest advertising model.

### **The Role of Diagnostics—Optimization**

Accountability is always a tricky problem when it comes to marketing activities, since the real world where professionals operate has always been an exceedingly messy multivariate place in which to perform research. The mere fact that brands exist in the world is proof that marketing works at a general level. However, determining whether marketing resources have been allocated efficiently and effectively has always been a problem.

For that reason, providing marketers with improvements that result in an incremental sense of control over even part of the complex marketing process will be rewarded generously by the marketplace. One of the biggest innovations in market research over the past half century has been the availability of retail scanner data. According to Moulton (1996), before the eighties, the average marketing budget was divided so that 43 percent of every marketing dollar was spent on advertising, while 57 percent was spent on consumer and trade promotion. By the early nineties, advertising’s share of the marketing dollar had shrunk to only 25 percent, and promotion had grown to 75 percent. What had changed, in the meantime, was a shift in the “balance of accountability.” With huge amounts of accurate retail sales data coming out of stores, it became easy to measure the short-term return on promotion while the effect of the advertising spending in the longer term became relatively less predictable and therefore riskier to use.

The desire to harness the creative power of advertising more effectively so as to attract those vanished client dollars has motivated the industry to embrace advertising tracking studies — witness the continuous tracking and marketing mix modeling services offered by Millward Brown — to justify advertising budgets. This put a spotlight back on the renewed need for pre-testing measurements to predict the performance of advertising prior to airing.

But meeting the client's need for accountability has long been a source of conflict within advertising agencies. In particular, most Creatives are skeptical of the value of copytesting. It is common for Creative Directors to caution us of the dangers inherent in “writing to the test,” where advertising is created based on a formula for what tests well rather than truly original work that re-defines a marketplace. They express concern, also, at the inefficiency of throwing away good creative ideas that are designed to operate differently than the researchers' models imply — the role of emotion in advertising is an example of the debates which arise from this longstanding issue. From an agency perspective, pre-testing research provides value only if it delivers an understanding of why a commercial scores the way it does and insights into how to improve an ad's performance — how to clean and polish the creative idea to a shining brightness.

One of the earliest agencies to embrace diagnostic copytesting was Leo Burnett. In the sixties, ad researchers Bill Wells and Clark Leavitt (in Wells, Leavitt, and McConville, 1971), and, later on, Fred and Mary Jane Schlinger (in Schlinger, 1979) developed a diagnostic pre-test for the agency called Communication Workshop. Extensive attitudinal rating statements were developed and factor-analyzed to provide a multidimensional profile of how a commercial was working in dimensions such as relevant news, believability, entertainment value and uniqueness. In addition, a series of open-ended questions were developed (based on the classic qualitative research funnel from general to specific) to provide insight into respondent reactions and message takeaway from the execution. A great deal of effort was also invested in developing complex coding schemes to better understand respondent reactions. For example, coding for product/narrative integration became a key source of insight into the well-brandedness of executions. The general way this form of research was positioned within the agency can be found in the policy that was upheld for many years that the research be conducted at agency expense and results be shared with Creatives for learnings, and shown to clients *only with the Creative Director's permission*. The test was designed only to be a tool for learning and optimization, not to be a report card.

Following Burnett's lead, many other agencies developed their own custom approaches to testing their own work. Key to the agency approach to copytesting is the belief that advertising operates in more than one way and that the methods that are effective in reaching consumers at one point in time may become less relevant as the consumer becomes vaccinated against various styles of advertising. This makes agencies suspicious of report card measures since the lessons from such systems always seem to point to one particular type of advertising approach. For example, “show the brand early and often” is advice often given by copytest systems based on DAR. Agencies also typically question the validity of a one-number approach to capture the complex workings of an ad.



Agency research departments were downsized in the eighties as a result of advertising's shrinking share of the marketing dollar and client price resistance to media inflation and consequent pressure on agency margins. Since that time, most agencies have gotten out of the business of testing their own ads. To most clients, this smacked too much of the fox watching the henhouse, anyway. Nevertheless, agencies tended to push hard for softer, more qualitative approaches to understanding their work early in the creative development process and to opt for focus groups rather than quantitative research as a mechanism for injecting the voice of the consumer into the creative development process.

According to the 4A's, the cost of commercial production doubled during the 1990's, with the average cost to produce a network quality television commercial rising from \$180,000 in 1989 to \$358,000 in 2002 (see [www.aaaa.org](http://www.aaaa.org)). This was twice the rate of inflation over the same period. At the same time, network audiences were shrinking. These factors all contributed to raising the risks associated with creating a television commercial but also shifted the interest of ad managers away from pure report-card testing systems like ARS and Ipsos-ASI and toward the hybrid systems combining validated performance measures with powerful diagnostics, systems like Millward Brown and Ameritest. The new need of the nineties was to squeeze every dollar's worth out of this increasingly expensive advertising film.

This shifted the value equation for the category, away from simply filtering creative ideas on the basis of one report card measure or another and put emphasis on gaining insights to improve the performance of individual executions. For some companies adopting this approach, the payback on research dollars has been considerable. In a recent review of the pre-testing work done by Ameritest over the course of a year by one of the major divisions which test large numbers of commercials, it was found that roughly forty percent of all ads being put on air had been revised or re-edited based on insights provided by quantitative diagnostic research, as shown in Exhibit 2.

Among copytesting suppliers, most diagnostic copytesting tended to follow the lead set by the ad agencies with some version of a test consisting of open-ended questions and attitudinal rating statements that bears more than a superficial resemblance to the Burnett model. But, for many practitioners, this purely verbal approach to describing the communication and response to advertising messaging seemed to be leaving important aspects of an ad's performance out of the picture. A clue to the limitations of the current research paradigm could be seen in the category descriptor: why call it "*copytesting*" when it's for *television*?

## The Search for the Subcontinents of the Mind

It has long been a belief of advertising professionals that much of the way advertising works operates below the surface. Not that we believe in the “subliminal advertising” techniques popularized by Vance Packard’s 1957 best seller, *The Hidden Persuaders*. Nor do we believe in advertising “magic.” But our own day-to-day experience is that much of the way that we interpret the world is subtext; the mind is continuously engaged in the search for deeper meaning, understanding the important gaps between what we say and what we do.

The old Chinese saying that “a picture is worth a thousand words” is completely untrue. There is information in every picture that we cannot put into words. There is something in the lovely sounds of our favorite music that we cannot verbalize—yet it moves us much more strongly than the literal meaning of the lyrics. This is the aesthetic information content of a television commercial. And the belief held by many advertising professionals is that traditional word-based research techniques — those protocols of open-ended questioning or ratings of semantic descriptions of the ad — somehow shortchange the best work.

For that reason, (and throughout the history of copytesting), there has been an interest in a range of nonverbal or physiological measures of advertising response. During the seventies, for example, Krugman (1997) published research on brain wave activity involved in viewing commercials. Other researchers experimented with galvanic skin response, voice pitch analysis and eye-tracking methodology — all non-conscious measures of a biologic response that researchers tried to correlate with commercial performance. Many of these early efforts met with disappointing results, usually because of the limitations of the technology that was used at the time. Recently, there has been a resurgence of interest in these types of measures, for example using MRI scanners to visualize brain activity during the commercial viewing experience (cf. Hall, 2002).

A different approach was developed by Ameritest in the nineties to probing response to advertising *film*. The Ameritest Picture Sorts is a method of deconstructing a viewer’s dynamic response to the film on multiple levels. The Flow of Attention, for example, measures how the eye *pre-consciously* filters the visual information in an ad and serves both as a gatekeeper for human consciousness and as an interactive search engine involved in the process of constructing perceptions of the brand. The focus of analysis is on understanding the role of film structure and syntax in creating those powerful film experiences that can provide the basis for the consumer’s emotional relationship with the brand. An example of a Flow of Attention graphic for one particular ad is shown in Exhibit 3.

More mainstream than the biological measures, the Ameritest Picture Sorts has been used around the world by major advertisers as diverse as IBM and Unilever in their standard pre-testing process. The strength of the diagnostic relationship between the different Ameritest “flow” measures and all the major report card measures produced by other pre-testing systems has been validated in a series of publications in the last few years (cf. Young & Robinson, 1989, 1992; Young & Kastenholtz, 2004; Young, 2002, 2004).

## Development of Moment-by-Moment Measures

One of the essential aspects of watching a television commercial is that it is a *temporal* experience — it's an unfolding of a sequence of ideas and emotions that takes place over time, in thirty seconds, or perhaps fifteen. From this research perspective, watching a television commercial is an event. Understanding the internal structure of that viewing event requires a very different approach to measurement.

The shift in analytic perspective from thinking of a commercial as the fundamental unit of measurement, a holistic and unified esthetic experience to be rated in its entirety, to thinking of it in terms as a structured *flow* of experience gave rise to experimentation with moment-by-moment systems in the early eighties. As an analogy, this transition in thinking is similar in some respects to the ways in which quantum physicists struggled to define the nature of light: is it a particle or is it a wave? Effectiveness measures such as recall or persuasion take a “particle” view of advertising; the moment-by-moment diagnostics implied a “wave” view of the advertising experience. And like the theoretical resolution of modern physics, advertising researchers are now beginning to acknowledge the fundamentally *dual* nature of advertising.

In the eighties, a number of firms commercialized moment-by-moment dial-a-meter response to television commercials. Viewfacts and, later, Millward Brown used a trace methodology to produce a one-dimensional response function based on rating liking or interest. While relatively little has been published to validate the relationships between these dial-a-meter systems and the general report card measures of attention, branding, or motivation, anecdotal evidence suggests this diagnostic technique can be a useful tool in the hands of a skilled interpreter.

A more recent development of the past few years has been the multidimensional Picture Sorts of the Ameritest system. Research from several experiments comparing the two approaches suggests that the Picture Sort flow graphs are measuring something fundamentally different from the information produced by the dial-a-meter system.

A key philosophical difference between the Picture Sorts and meter ratings approach is that, with the Picture Sorts, quantitative measurement is a function of the rate of information flow (or visual complexity) of an ad whereas dial-a-meter approaches are based on mechanical clock time. The Picture Sorts approach shifts the frame of measurement to the flow of subjective time in the audience's experience of the advertising. In addition, measurement is made in three dimensions, not one: the Flow of Attention, which measures how the parts of the ad are processed on a cognitive level; the Flow of Emotion, which measures how the emotions being produced by the ad can be processed in four different ways to achieve different dramatic effects (as shown in Exhibit 4); and the Flow of Meaning, which quantifies how the emotions created by the ad are harnessed to illuminate a brand's core values, what we call the “working emotion” in an ad. The latter type of emotion is also what Aristotle called “aesthetic emotion,” or emotion that carries with it a powerful meaning into the mind of the consumer.

## **The Future: Seven Trends**

This, then, is basically the state-of-the-art, circa the beginning of the twenty-first century. Looking backward for the past fifty years, we see that the art and science of measuring advertising has undergone a considerable amount of debate and change in our thinking and in our business practices. Looking around at the forces of change operating in our world today, there is no reason to think that the rate of change will slow down. The following are seven trends that will be shaping the way we do business in the future.

### **1. The emergence of global research standards for advertising global brands.**

Increasingly, multi-nationals are focusing on the need to build global brands, and for their brands to speak with one voice around the world. This calls for global advertising campaigns that will be increasingly visual in style. Deploying global research systems to provide a standard way to measure advertising performance from one region to another and for providing tools for making transparent how different cultural factors affect advertising response will become an important management focus for managing ad spending in the global marketplace.

### **2. There will be more advertising measurement, not less.**

Advertising is becoming more expensive and the range of executional options have become so diverse that more control over the process is being demanded by major clients today. Procurement departments, in particular, under the banner of accountability, are challenging advertising agencies and research companies to provide more proof of value to justify ad budgets. This will drive growth in this important sector of research.

### **3. Most copytesting will move on to the internet.**

In an age of rapid-response marketing, the emphasis is on speed of decision-making. The internet is the obvious choice for shortening the time involved in the research step of the creative development cycle. Many suppliers like ourselves have already begun migrating their advertising research online (for both television and print testing) and economic pressure will probably force the majority of testing online in the near future.

### **4. The new value proposition will be filtering plus optimization.**

For the foreseeable future, the cost of advertising executions will continue to go up. To manage that cost, managers will be increasingly concerned to make sure not only that they air their strongest ideas, but that they don't spend half their advertising budgets on *average* ideas. Ad managers will be looking for every opportunity to make executions work harder and research systems will outperform this growing category if they can *validate the power of their diagnostics*, providing proof that they actually help make ads more effective.

**5. Ad research will move beyond semantics—putting a new emphasis on non-verbal measurement.**

Both the forces of globalization and the evolution of rich, multi-sensory media environments will continue to challenge researchers to think beyond the boundaries of language and semantics in understanding how advertising builds brand image. Leading ad researchers of today, such as Gerald Zaltman of Harvard University (cf. Zaltman, 2003) have pointed out that new learnings from neuroscientists over the last few years have challenged us to develop new methods and approaches to understanding the hidden responses of the mind. Currently the Advertising Research Foundation and the 4A's in the US are sponsoring a major initiative on understanding the role of emotion in television advertising — which will be reported out in series of events over the next year.

**6. New heuristic models will be built to help managers make ad decisions in a world increasingly confused by media fragmentation.**

As the world of media becomes increasingly fragmented and media choices proliferate, the need for research is to *simplify* the decision-making process for advertising managers. This calls for new *heuristics* that describe how different media work, e.g., television versus print. These heuristics are necessary to provide a common measurement framework so that advertising managers who are trying to allocate budgets across television or print or the internet can compare the relative strengths of the television executions to the print execution to the internet ad.

**7. New mathematical approaches will be developed to model advertising effects.**

Currently, researchers working in the field of complexity and chaos theory have been revolutionizing approaches to studying complex, messy problems such as describing the seemingly random behavior of financial markets or in describing the complex flows of industrial supply chains. A world center for this activity, close to home for me, is the Santa Fe Institute in New Mexico (see [www.santafe.edu](http://www.santafe.edu)). This is an example of new dynamic, non-linear approaches to building computer models that move well beyond the predictive power of the traditional linear approach of regression modeling. Up to now, few efforts have been made by advertising researchers to apply these mathematical techniques to advertising measurement — the very definition of a non-linear problem. But it's only a matter of time.

## References

- Arnold, S. J., & Bird, J. R. The Day After Recall Test of Advertising Effectiveness: A Discussion of the Issues. In Current Issues and Research in Advertising, 1982, 59-68.
- Blair, M. H. An Empirical Investigation of Advertising Wearin and Wearout. Journal of Advertising Research, 40 (November/December 2000).
- Brown, G. Insights into Campaign Wear-in and Wear-out from Continuous Tracking. Paper presented at the ARF Advertising Tracking Studies Workshop, New York, 1984.
- Brown, G. Campaign Tracking: New Learning on Copy Evaluation and Wearout. Paper presented at the Fourth Annual ARF Copy Research Workshop, New York, 1987.
- Hall, B. F. A New Model for Measuring Advertising Effectiveness. Journal of Advertising Research, March/April 2002, 23-31.
- Honomichl, J. J. Honomichl on Marketing Research. Lincolnwood, IL: NTC Business Books, 1986.
- Kastenholz, J., Kerr, G., & Young, C. Focus and Fit: Advertising and Branding Join Forces to Create a Star. Marketing Research, Spring 2004, 16-21.
- Kastenholz, J., Young, C., & Kerr, G. Does Day-After Recall Testing Produce Vanilla Advertising? Admap, June 2004, 34-36.
- Krugman, H. Memory Without Recall, Exposure Without Perception. Journal of Advertising Research, July/August, 1977.
- Lodish, L. M., Abraham, M., Kalmenson, S., Livelsberger, J., Lubetkin, B., Richardson, B., & Stevens, M. E. How TV Advertising Works: A Meta-Analysis of 389 Real World Split Cable TV Advertising Experiments. Journal of Marketing Research, May 1995, 125-139.
- Moult, W. H. Selective Success Amid Chaos: Advertising in the 1990s. Paper presented at Esomar, 1995. In "Selected Papers and Presentations: 1990-1996," Stamford, CT.: ASI Market Research, Inc., 1996.
- Ostlund, L. E., Clancy, K. J., and Sapra, R. Inertia in Copy Research. Journal of Advertising Research, 1980, 20(1), 17-23.
- Ross, H. Recall vs. Persuasion: An Answer. Journal of Advertising Research, 1982, 22(1), 13-16.
- Schlinger, M. J. A Profile of Responses to Commercials. Journal of Advertising Research, 1979, 19(2), 37-46.

Wells, W. D., Leavitt, C., and McConville, M. A Reaction Profile for TV Commercials. Journal of Advertising Research, 11(6), December 1971, 11-17.

Young, C. E. Researcher as Teacher: A Heuristic Model for Pre-Testing TV Commercials. Quirk's Marketing Research Review, 2001, 23-27.

Young, C. Brain Waves, Picture Sorts, and Branding Moments. Journal of Advertising Research, 42(4), 2002, 42-53.

Young, C. The Eye Is Not A Camera. Quirk's Marketing Research Review, 2003, 58-63.

Young, C. E. Capturing the Flow of Emotion in Television Commercials: A New Approach. Journal of Advertising Research, June 2004, 202-209.

Young, C., & Kastenholz, J. Emotion in TV Ads. Admap, 2004, 40-42.

Young, C. E., & Robinson, M. Video Rhythms and Recall. Journal of Advertising Research, June/July 1989, 22-25.

Young, C. E., & Robinson, M. Visual Connectedness and Persuasion. Journal of Advertising Research, March/April 1992, 51-59.

Zaltman, G. (2003), How Customers Think: Essential Insights Into the Mind of the Market, Boston, MA: Harvard Business School Press.