

ONE SIZE FITS ALL

A global model for diagnostic pre-testing of TV commercials

Chuck E. Young

The paper describes the international validation of a heuristic model for pre-testing TV commercials. The validation is based on the model's internal measures rather than external comparisons such as pre- and post-test validations. The model states that all effective advertising, anywhere in the world, must accomplish three goals: attract an audience, clearly link the ad to the brand in the viewer's mind, and motivate the viewer to take action. The model clearly defines key measures that determine Attention, Brand Linkage and Motivation scores. This paper reports data obtained from over one hundred thirty (130) pre-tests conducted in the United States, Europe and Asia Pacific for multiple clients.

INTRODUCTION

The development of multi-national advertising campaigns requires sensitivity to the diverse cultural lenses through which different audiences in our global community view the iconography of an advertisement. Yet it also requires the operating assumption that on at least some levels advertising works in the same way around the world. The implicit argument is that in the end all advertising works by making a connection with the human mind.

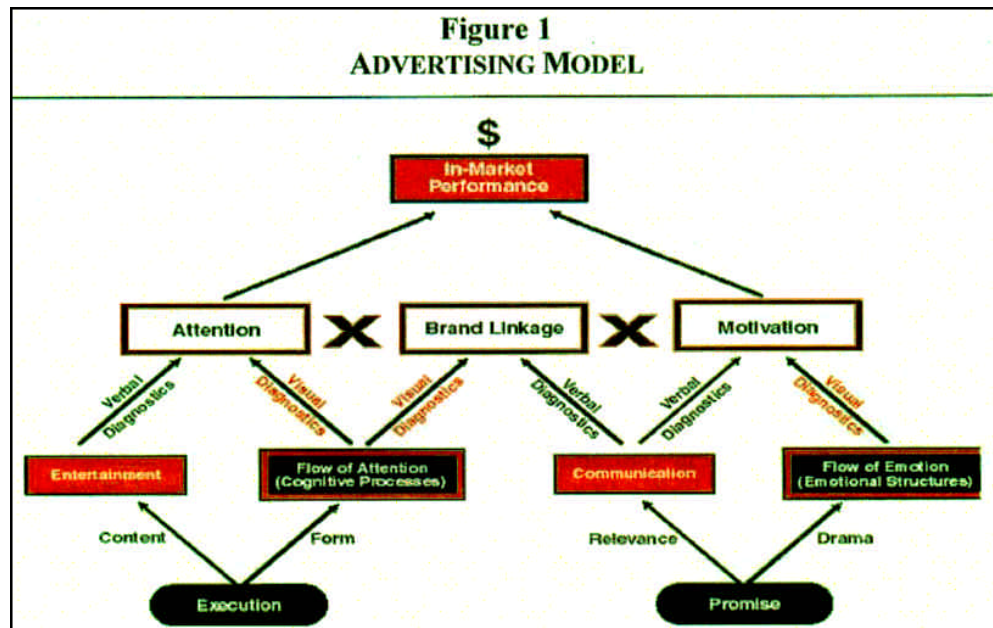
For advertising managers, charged with the responsibility of directing advertising for a global brand in a multi-cultural world of increasing change and clutter, there is an important need for a simple and coherent framework for thinking and communicating about advertising in a way that crosses the cultural divides that separate managers in one country from those in another. Indeed, client and agency are sometimes as far apart in terms of communication as any two countries. A similar divide separates traditional market researchers from account planners (Baskin and Coburn, 2001).

VALIDATING DIAGNOSTIC RELATIONSHIPS

Even within a single country the issue of validating commercial pre-testing performance measures against actual in-market sales effectiveness is a difficult and ongoing problem (Haley and Baldinger, 1991). Generalizing the problem across national boundaries is a much greater challenge to researchers given the confounding effect of variable market conditions, differences in brand development, and other marketing and cultural factors. This larger problem is well beyond the scope of this paper, instead, our focus is on the issue of determining the quantitative relationship between a variety of diagnostic constructs that are commonly used to describe the experience that a viewer has of a television commercial—variables which can be measured within the course of a single respondent interview.

A HEURISTIC MODEL FOR DIAGNOSTIC PRE-TESTING

A “heuristic” model is one that is primarily used for “teaching purposes”—in our case, to help advertising decision-makers understand important relationships in the complex problem of advertising measurement so as to bring clarity with regard to what the research is telling them. Simplicity, therefore, is prized at the expense of showing all the relationships that might be included if we were interested in building a predictive model. This clarity is of paramount importance, we have found, if research is to be used effectively for decision making by non-researchers.



This heuristic model which shows the primary relationships between different commercial variables provides clients with a “mental road map” to major pre-testing measures.

In the Ameritest™ heuristic model, shown in figure 1, information is arranged in a hierarchy of importance. At the top of the model is what pre-testing is supposed to predict: in-market sales results. One level down are the evaluative measures that provide the report card portion of the analysis. Two levels down are the diagnostic measures that are correlated with, and therefore help to explain, the evaluative measures above.

The arrows in the model highlight the primary relationships between the different variables measured and hence provide a “road map” for interpreting a complex set of research data.

Essentially, the heuristic model says that for any commercial to be effective it must accomplish three distinct things:

1. It must get noticed and attract an audience.
2. The audience must know who is sending the advertising message.
3. Once the commercial has the audience’s attention it must sell them something — i.e., motivate sales in the short run or at least create a positive pre-disposition for sales in the long run.

Other variables are important only insofar as they help to explain the variables of Attention, Brand Linkage and Motivation. For example,

Entertainment value is not important in and of itself but only because it is an important predictor of Attention.

The model shows us that Attention is a function of two primary aspects of the *commercial execution*:

1. Entertainment value: Does the execution entertain or reward viewers with an enjoyable experience in return for the 30 seconds they are asked to spend with the client's message?
2. Flow of Attention™: This is a moment by moment measure of cognitive processing that deals with the advertising viewed as a piece of film. Is the execution a well-edited piece of film that captures and maintains the viewer's attention over time, focusing thoughts and feelings on the important ideas and images in the commercial at a pace the viewer can easily keep up with?

The model also shows us that Motivation is a function of how well the *strategic promise* is being delivered:

1. Communication: Does the commercial deliver a relevant, newsworthy and believable message to the viewers?
2. Flow of Emotion™: How much emotion does your commercial tap into with the magic power of film to make your brand's promise seem "larger than life" and even more compelling than the rational concept? In other words, the key to motivation is communicating a relevant idea in a dramatic way.

It is more than an academic exercise to build heuristic models. If teaching clients to use research to make smarter decisions about advertising gives them a competitive advantage in the marketplace, then improving your models over time is a way of sustaining that competitive advantage.

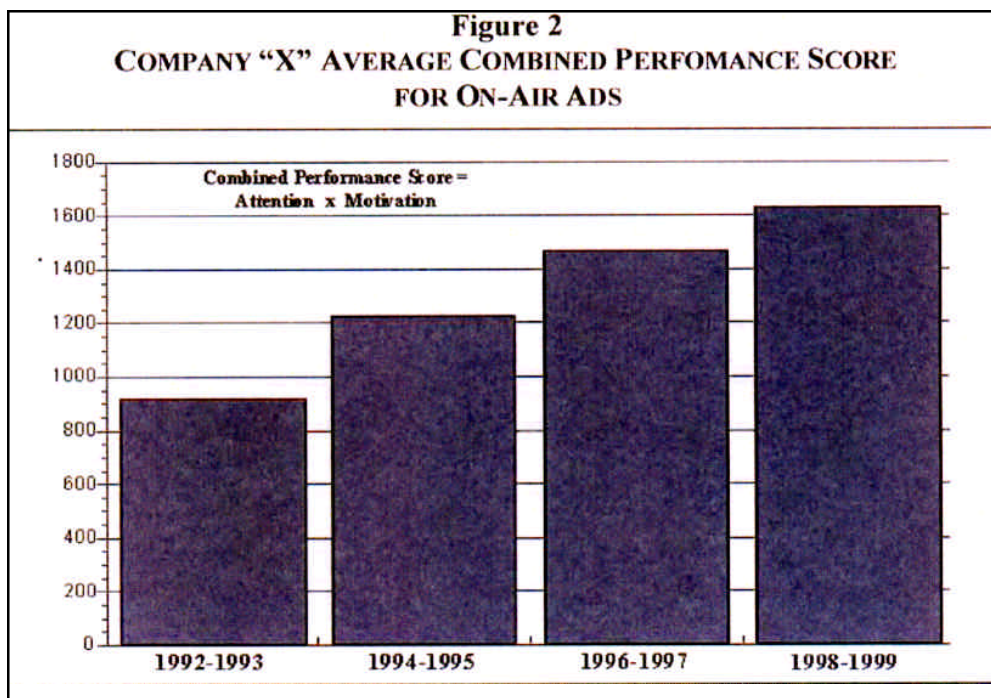
A LEARNING EFFECT

Pre-testing research can be used to improve advertising performance in one of two ways: 1) it can be used as a "filter", to screen out weak advertising ideas before they are put on air, or 2) it can be used as input to a process of "optimization", to help improve the performance of good advertising ideas that are not always perfectly executed.

The first approach, which simply uses research as a "report card" on the advertising, operates in a straight-forward, albeit mechanical way: to raise the performance of the advertising that is put on air, management simply establishes a hurdle rate for a "go/no go" decision about performance on

key measures. While this quality control approach would logically raise the average *level* for on-air advertising, it would not necessarily lead to an ongoing improvement in advertising performance over time.

The second approach uses research as an input to a “learning process”, with the goal of providing insights and understanding as to the reasons why executions are performing the way they do. The lessons provided by this kind of diagnostic research can be actionable in one of two ways: 1) flawed executions can be re-edited for improved performance; and 2) by providing a feed-back loop for understanding why executions are working as they do, new commercial “pool-outs” for an ongoing campaign or even new campaigns can be developed that build on a growing base of knowledge. Empirical evidence that actual learning by the ad team is taking place over time should be seen in an on-going improvement in commercial performance.



An example using a large base of test ads demonstrates the competitive advantage gained from utilizing a “learning” system.

Base: 200 commercials since 1997.

We have evidence that such a “learning effect” has taken place with a company that, about ten years ago, adopted the heuristic model reported above as part of a process for managing advertising development, as shown in figure 2. Over that ten-year period approximately 200 commercials were test-

ed in the United States in several closely related product categories. During this period the average performance of ads being put on air, measured by both Attention-getting power and Motivation, has increased steadily. At the same time, the average performance of a substantial number of competitive ads that the company has also tested — for learning purposes — has not changed in a systematic way. Additionally, sales for the advertised brands have generally increased along with advertising performance.

In other words, the advertising managers at this company appear to be getting smarter over time. We immodestly suggest a systematic learning process based on diagnostic research, and the heuristic model for interpreting the research data, have made a contribution to this result

To borrow from Peter Senge's book on learning organizations. *The Fifth Discipline*, "The effectiveness of a leader is related to the continual improvement of the leader's mental models" (1994). The challenge to companies interested in leading the way toward building global brands, therefore, is to test the assumption that the mental models that we have developed to explain how advertising works within one country work as well in other countries.

VALIDATING THE HEURISTIC MODEL: A META-ANALYSIS OF INTERNATIONAL TELEVISION COMMERCIAL TESTS

This paper reports data obtained from a meta-analysis of over one hundred and thirty (130) pre-tests conducted for several clients in the normal course of business. While for reasons of confidentiality actual test scores cannot be reported, it is important for purposes of face validity to note that the underlying information upon which this analysis is based, was used for actual business decision making and was felt to be in general accord with the judgment of the experienced advertising professionals making these decisions.

The Sample of Television Commercial Tests

The sample used in this study consists of finished television commercials conducted by several multi-national advertisers in the United States, European and Asia-Pacific markets during the past three years. Specifically, 90 ads were tested in the United States, 23 were tested in Europe and 19 were tested in Asia-Pacific. The target for these ads were both business-to-business and consumer audiences in approximately the same proportion across the three regions.

Some, but not all, of the commercials were tested in all three regions and it should be noted that the same ad did not always perform equally well in all



three markets. This confirms the importance of understanding cultural differences when developing a multi-national advertising campaign. Some of the commercials were developed and tested locally while others were developed in the “headquarters” country and exported for local use with appropriate translations or subtitles — our sample is not large enough, however, to draw any conclusions about these differences. In general, the sample of ads in this study includes both strong and weak performers.

The Research Methodology

The methodology of the test was the same in all three regions, though local, in-country research suppliers handled data collection. Each commercial was tested among 150 target respondents in a personal interview lasting roughly 25 minutes. Each respondent was first exposed to the test ad in a clutter environment along with four other commercials competing for their attention. Measures of Attention-getting power (or “breakthrough”), and Brand Linkage were obtained. Next, the respondent was shown the test ad by itself and measures of motivation, communication and other diagnostics were collected.

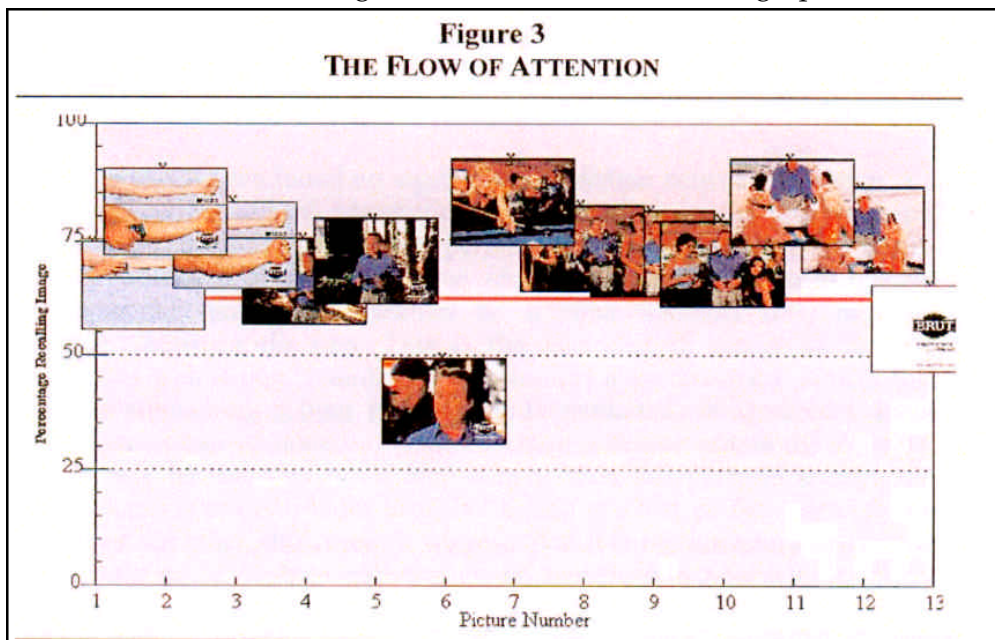
The measure of Attention is obtained after the clutter exposure by asking the respondent “Which of the ads you just saw did you find interesting?” Note that this is not a measure of memory or recall, simply a measure of audience interest. This measure is preferred for several reasons. First, it is a very discriminating measure when asked even a short time after the exposure to a clutter reel. Second, we find that it lines up well with our client’s in-market tracking study results. Third, in a soon-to-be-published paper (Young and Cohen) we will be reporting that it correlates very strongly with ads that have won major creative award shows — contests which we feel inherently reward “break-through” creativity — and gives the measure face validity for agency creative directors.

Brand Linkage is measured by coding whether or not the brand name was used when a respondent reported that a particular ad was one of the interesting ones. In other words, our measure of Brand Linkage assesses whether or not the brand name is being used as the “retrieval hook” for describing the ad. It is, in that sense, not unlike a top-of-mind or unaided measure of brand awareness that would be used in a conventional tracking study.

Motivation for the ads reported on here is a simple five-point probability intention to buy/do business with scale (e.g. “Definitely will/Probably will...”) such as might be used in a conventional concept test. Verbal diagnostic ratings are collected using a six-point strongly agree to strongly disagree scale.

Non-verbal measures are based on the Ameritest Picture Sorts™ method described by Young and Robinson (1987). In brief, during this last part of the interview, respondents are either handed a shuffled deck of photographs or shown the images on screen in a CAPI interview. These images, typically 10 - 30 pictures for a 30-second commercial, are taken from the commercial. Because the images are based on the commercial iconography itself, this deck of photos provides a *natural vocabulary* for interrogating respondents about their visual experience of the advertising — without using words.

Armed with this new vocabulary, respondents first sort the pictures on the basis of recall: those they remember seeing in the ad and those they do not. Next, respondents sort the pictures based on emotional response — by sorting them into six groups based on their feelings as they watched the commercial, from strong positive to strong negative response. In general, the two sorts produce quite different patterns of data. Data from the first sort is used to generate a Flow of Attention™ graph (see figure 3), while data from the second sort is used to generate a Flow of Emotion™ graph.



The height of each picture shows the percentage of the audience recalling that picture. According to the model, this technique provides diagnostic insights into Attention and Brand Linkage.

MULTIPLE DIMENSIONS OF “REPORT CARD” PERFORMANCE

A correlation analysis was performed on the samples of commercials in each of the three regions for the Attention, Brand Linkage and Motivation measures. Table 1 reports the results.

Table 1

Performance Measures	INDEPENDENT PERFORMANCE MEASURES								
	US			Europe			Asia-Pacific		
	Attention	Brand Linkage	Motivation	Attention	Brand Linkage	Motivation	Attention	Brand Linkage	Motivation
Attention	1.0	-.09* (.069)**	.14 (.172)	1.0	-.44 (.034)	.14 (.571)	1.0	.02 (.930)	-.14 (.635)
Brand Linkage	-.19 (.069)	1.0	.14 (.186)	-.44 (.034)	1.0	-.29 (.230)	.02 (.930)	1.0	-.13 (.648)
Motivation	.14 (.172)	.14 (.186)	1.0	.14 (.57)	-.29 (.230)	1.0	-.14 (.635)	-.13 (.648)	1.0
	N = (92 ads)			(23 ads)			(19 ads)		

* Pearson Correlation; ** 2-tailed Significance

In all three regions we found no significant correlation between Attention and Motivation, confirming the assumption of the heuristic model that these are independent measures of commercial performance. This is an important result. In our experience it appears that some advertising creatives seem to operate with a mental model that assumes an implicit trade-off (i.e., negative correlation) between the two. This is the idea that if one is to develop commercials with strong break-through power one must down-play content that delivers a compelling selling proposition. In particular, disagreement about this premise seems to underlie much advertising debate across the Atlantic. While it may be true that to develop commercials that perform strongly on both measures is more difficult than developing ads that perform strongly on only one or the other, this research suggests that it is not necessary to give up one to get the other. Such a creative trade-off appears to be a false trade-off.

Brand Linkage also appears to be independent of the other two evaluative measures. There is no significant correlation between Brand Linkage and Motivation in any of the three regions. There is a weak, negative correlation between brand linkage and attention in the United States and Europe, but not in Asia-Pacific. Again, we suspect the slight negative correlation between attention and brand linkage is the result of an implicit trade-off some creatives may be making; for instance, there certainly appears to be a growing trend to withhold brand identification until the end of commercials, a creative choice which appears to be based in part on the assumption that to get the attention of an audience one must get under the psychological defenses that consumers have developed against advertising by having the ad not look like an ad for as long as possible.

These three independent variables, then, form a multi-dimensional approach to evaluating advertising performance. While it would certainly be possible to combine the three into a single numerical index, our experience suggests that doing so produces a "black box" effect and works at cross-purposes to the learning nature of diagnostic research.

The Communication factor includes items related to “news value”, “message importance” or “relevance”, “believability” and, the opposite of communication, “confusion”. Again, these items are highly correlated with the Motivation measure in all three regions, and are either uncorrelated or much less correlated with the other two evaluative measures. This, too, is consistent with the model.

In general, the Brand Linkage measure is not strongly explained by these verbal rating statements — though it does seem to be somewhat related to ads that convey relevant news,

Overall, it appears that the primary verbal diagnostic relationships shown in the heuristic model are good across all three geographic regions.

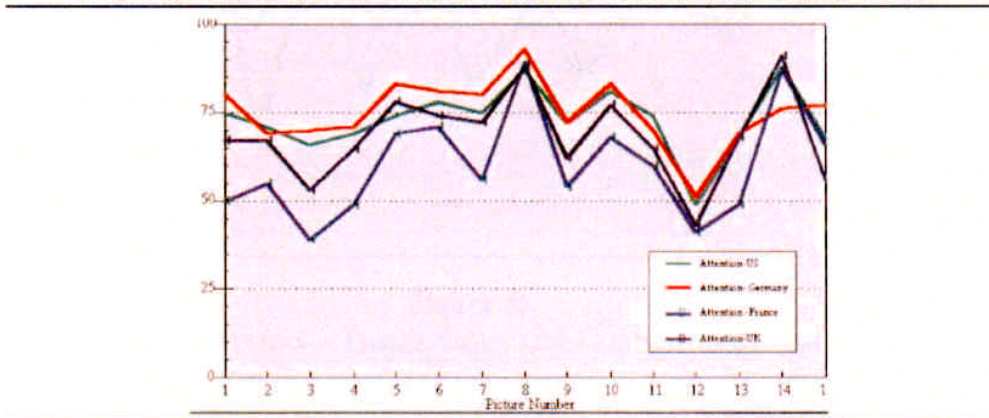
EXPLAINING PERFORMANCE SCORES WITH NON-VERBAL MEASURES

Visuals cross language boundaries and are the key to global advertising campaigns. Yet the old adage that a picture is worth a thousand words is simply untrue in every piece of communication. There is esthetic information that cannot be put into words. That is the message of the French musicologist Abraham Moles in his seminal work *Information Theory and Esthetic Perception* (1966).

The Flow of Attention and Flow of Emotion are elements of the model that help us to understand how a commercial is working moment by moment, from the standpoint of the aesthetic effects of film. From a heuristic or teaching standpoint, the Flow of Attention graph helps the advertiser make the paradigm shift from thinking of the human eye as a recording device, like a camera, to thinking of it more as a computerized search engine that actively sorts through information. Selective perception is the filter that alters an advertising message from what the advertiser intended to what the viewer actually understood.

Just as the Flow of Attention helps us to better understand Attention and Brand Linkage scores (Young and Robinson, 1987, 1989), a Flow of Emotion graph helps to explain Motivation scores. Both graphs are analyzed in structural terms, with respect to the form and executional content of the wave functions that are produced by the data, not just in terms of the absolute levels of response. Consequently, for purposes of this validation exercise, rigorous quantitative confirmation is still in progress. Results that we are reporting here are preliminary — but encouraging.

Figure 4a
FLOW OF ATTENTION – FOUR COUNTRY COMPARISON FOR AD X



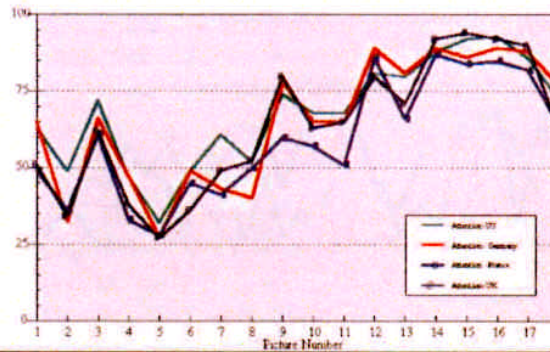
This pattern of attention to the visuals is similar across all four countries. Comparable results have been found for tests conducted in Japan and China.

Figures 4 and 5 show the Flow graphs for several executions that were re-tested in several countries. As you can see, the patterns that are being produced are quite similar across countries and these results are typical of our experience so far. As in the United States, visual analysis of commercials tested in Europe and Asia-Pacific has contributed insights, complementing traditional verbal methods, for explaining Attention, Brand Linkage and Motivation performance scores.

IMPLICATIONS FOR ADVERTISERS

This research has helped demonstrate the proposition that whether or not television commercials are aired in the United States, Europe or the Asia Pacific region, they can be consistently evaluated with the independent variables of Attention-getting power. Brand Linkage and Motivation — constructs which, on the face of it, would seem to be essential to in-market sales effectiveness. Moreover, regardless of which region you are talking about, the kinds of executions that we have found get the attention of the audience are those that reward the viewer for the thirty seconds of time you are asking them to give you — by entertaining them, showing them something they like or find imaginative and that is worth seeing more than once. And, regardless of country, the way to motivate that audience is by communicating a relevant, believable and newsworthy message that is not confusing to your audience. Importantly, there is, in general, no inherent trade-off between the ability of a commercial to get noticed and the ability of that ad to sell a brand — the ads that your audience really likes manage to do both.

Figure 4b
FLOW OF ATTENTION – FOUR COUNTRY COMPARISON FOR AD V



Again, the pattern of response to the visuals is similar across all four countries.

Figure 4c
FLOW OF ATTENTION

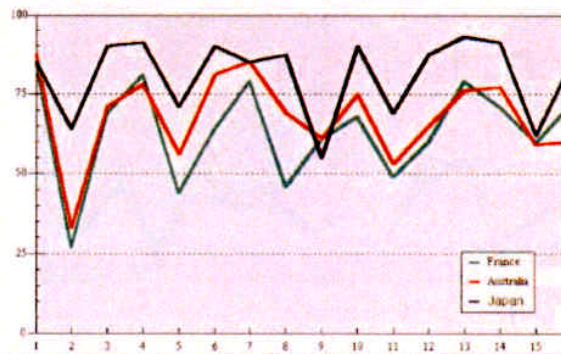
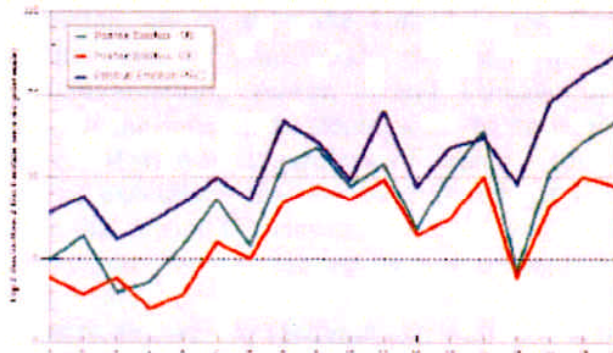
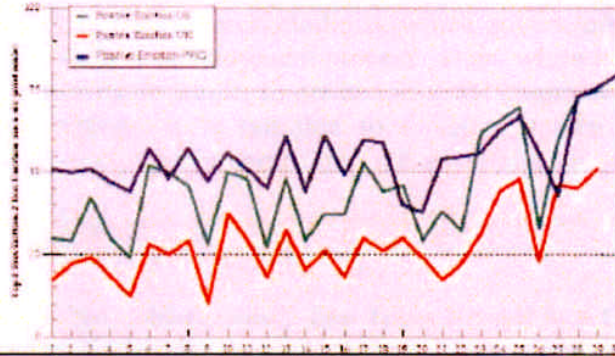


Figure 5a
FLOW OF EMOTION – THREE COUNTRY COMPARISON FOR AD X



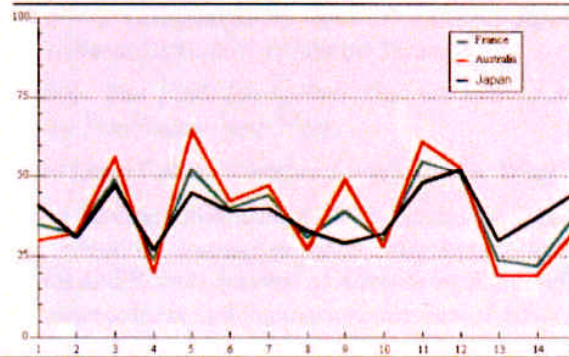
This pattern of attention to the visuals is similar across all three countries. Comparable results have been found for tests conducted in Japan and China.

Figure 5b
FLOW OF EMOTION – THREE COUNTRY COMPARISON FOR AD V



Again, the pattern of response to the visuals is similar across all three countries.

Figure 5c
FLOW OF EMOTION



IMPLICATIONS FOR ADVERTISERS

Importantly, by validating a simple heuristic model with international data we have shown that it is not necessary to re-invent, country-by-country, advertising theory or the research methods, which advertisers can use to bring discipline to the creative development process. Thus while it may not be easy, or sometimes not even desirable, to create universal commercial executions for building global brands, it is possible to develop standards for measuring quality for managers to use in controlling the advertising process with a global perspective.

REFERENCES

Baskin, Merry and Neil Coburn. (2001). Two Tribes Divided by a Common Language? The True Nature of the Divide Between Account Planners and Market Researchers. *Market Research Society Conferences*, 21–34, March.

Haley, Russell and Allan Baldinger. (1991). The ARF Copy Research Validity Project. *Journal of Advertising Research*, 31 (March/April).

Moles, Abraham. (1966). (Translated by Joel E. Cohen.) *Information Theory and Esthetic Perception*, Urbana: University of Illinois Press.

Senge, Peter M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*, Current Doubleday, New York.

Young, Charles E. and Larry Cohen. Creative Award Shows: What are They Telling Us?

Young, Charles E. and Michael Robinson. (1987). Guideline: Tracking the Commercial Viewer's Wandering Attention. *Journal of Advertising Research*, 27 (June/July), 15–22. (1989). Video Rhythms and Recall, *Journal of Advertising Research*, 29 (June/July), 22–25. (1992). Visual Connectedness and Persuasion, *Journal of Advertising Research*.

THE AUTHOR

Charles E. Young is CEO and Founder, CY Research, Inc. d/b/a Ameritest, United States.

