

Does day-after recall testing produce vanilla advertising?

John Kastenholz, Unilever, Chuck Young, Ameritest, and Graham Kerr, Millward Brown, discuss the evidence that recall testing favours boring ads

IN TERMS OF MARKET share, day-after recall testing is still the leading pre-test methodology in the US for measuring TV commercial breakthrough. Yet many creative directors have argued over the years that day-after recall frequently produces results that are counterintuitive to them. This leads many to cynically conclude that pre-testing research undermines their best attempts to produce breakthrough creative ideas. This is particularly true of commercials that are designed to break through media clutter by being entertaining or emotional and not just to communicate product news.

In today's advertising climate, with pressure for both advertising accountability and breakthrough creative ideas, the need has never been greater for understanding how to measure advertising effectiveness. Unilever has made a considerable investment in research dollars over the last year comparing different commercial measurement approaches in order to improve its advertising research process. One of the outcomes of that research is new insight into the executional variables that relate to attention and recall. The purpose of this article is to share some of our findings in order to stimulate further discussion and innovation in this important area of advertising research.

The database

Over a one-year period, Unilever triple-tested 60 television commercials in both the Ameritest, the Ipsos-ASI, and

TABLE 2

Audio-visual relationships to 'breakthrough' measures

	Ipsos ASI recall	Ipsos ASI attention	Ameritest attention	Millward Brown attention
Verbal diagnostics				
Early audio category mention	0.11	-0.23	-0.35**	-0.26*
Early audio brand mention	0.18	-0.23*	-0.37**	-0.35
Total audio brand mentions	0.40**	0.30**	0.05	0.01
Visual diagnostics				
Total number of peaks	0.03	0.23	0.35**	0.42**
Number of executional peaks	-0.02	0.27*	0.40**	0.52**
Number of product peaks	0.27*	0.12	-0.08	-0.07

Note: * > 95% confidence ** > 99% confidence

Millward Brown pre-testing systems, a multi-million-dollar investment in research. Since these are standard, commercially available tests, descriptions of these methodologies can be obtained by visiting the companies' websites and will not be described here.

These commercials were produced for home and personal care product categories. Approximately half the executions were finished film, while the other half were tested in an animatic or rough stage of production. Nearly half the advertising executions were for new products – in general, line extensions of Unilever masterbrands – and half were for established brands. The executions represent the work of a number of major advertising agencies and creative teams and thus express a range of creative, approaches. But, in terms of corporate advertising philosophy, this set of commercials tips more towards the emotional end rather than the rational end of the advertising continuum.

We were asked to analyse this unique database of commercial performance scores in order to understand the strengths and weaknesses of three different measurement philosophies.

Recall and attention

Most copy-testing systems report some kind of measure of 'breakthrough', either recall or attention, as an indicator of how

efficiently a commercial execution will leverage a given level of media weight to capture a wide audience for the advertiser's message. The correlation between recall and the attention measures reported by these three systems is shown in Table 1. Observing that they are in fact uncorrelated measures, we arrive at our first conclusion: recall and attention cannot both be measuring 'breakthrough' power – they are measuring fundamentally different aspects of commercial performance.

We also notice that even though they all call it 'attention', the three systems are measuring that idea differently as well. The ASI measure of attention, which is based on a respondent's ability to recognise an ad from a verbal description of a commercial, has a positive correlation with recall – probably because both ASI measures are fundamentally derived from the attempt to describe TV commercials in words. In contrast, the Ameritest measure, which is based on whether or not a commercial is found to be interesting in the context of a clutter reel exposure, is completely independent of the recall measure. Finally, the Millward Brown measure, which is based on a respondent's reported active enjoyment of a commercial, is, in fact, negatively correlated with recall. In general, as we will see in the analysis below, the Ameritest and Millward Brown measures of attention are

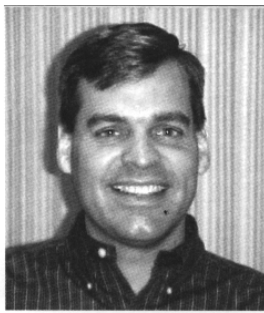
TABLE 1

'Breakthrough' measures

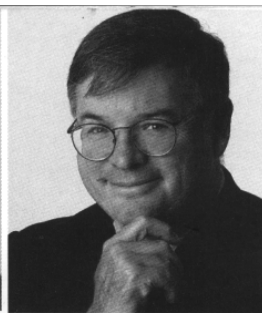
Correlation between recall and attention measures of the three pre-testing systems

	Ipsos ASI recall
Ipsos ASI Attention	0.67**
Ameritest® Attention	0.09
Millward Brown Attention	-0.28*

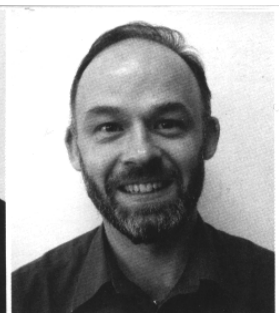
Note: * > 95% confidence ** > 99% confidence



John Kastenholtz is vice president of consumer and market insight, at Unilever home and personal care in North America.



Chuck Young is founder and CEO of Ameritest, an international advertising and brand research company.



Graham Kerr is executive vice president of the eastern division of Millward Brown in the US.

more similar to each other and different from the ASI measure of attention.

Diagnostic relationships to recall and attention

In order to see past the labels, we need to examine diagnostic variables that explain what it is we are really measuring when we talk about attention and recall. First, we looked at verbal, and then non-verbal, diagnostics for insights into these different performance measures, as shown in Table 2.

For each of the commercials in our sample we counted the number of seconds from the beginning of the ad until the product category and the brand was first mentioned. We also counted the number of times the brand was mentioned in the ad. The most important of these audio variables for explaining recall is repetition of the brand name. Repeating the brand name is correlated with the ASI measure of attention, but it is unrelated to the Ameritest and Millward Brown measures of attention.

Providing early audio cues for either the product category or the brand was negatively correlated with all three measures of attention. This finding suggests that there is merit to the contemporary creative 'under-the-radar' approach to capturing audience attention by withholding category and brand cues until later in the commercial.

Indeed, as shown in Table 3, early category or brand cues tend to be associated with commercials in our sample of 60 ads that, on average, are rated by consumers to be less entertaining, interesting, involving and unique, and that are more boring and ordinary. Moreover, multiple brand name mention, which is one of the keys to recall, is also associated with boring and ordinary advertising.

The visual analysis is based on the Ameritest Flow of Attention patterns generated for the commercials in this sample. This flow measure of attention is based on the insight that the human eye should not be thought of as a simple recording device like a camera, but rather should be thought of as an intelligent gatekeeper of perception that pre-consciously filters visual stimuli as part of the complex mental process of constructing visual perception. The technique is based on sorting a sample of visual frames taken from a commercial into one of two categories: those they recognise and those they do not recognise from a forced viewing of the commercial.

An example of a Flow of Attention® graph for Unilever's Degree deodorant is shown in Figure 1. Of interest here are those frames that we classify as peak moments or focal points of visual attention in the commercial. Peak moments are defined as local maxima in the flow curve – that is, images that are relatively

higher than those in the neighbourhood – and are not defined relative to a norm or some absolute level of recognition. (Also, by convention, the first image in a flow graph is not counted as a peak.) In the example, there are three peaks.

Peaks may be classified into two categories, based on the type of visual information contained in the peak. The first type is product-related information – by that we mean package shots, product demonstrations, product-in-use shots, specific product claims, or information about support points such as ingredients, or simply the brand name. All other peaks are, by default, classified as executional, which is our advertising adaptation of the more technical description of 'esthetic information' provided in the French musicologist Abraham Moles' classic work, *Information Theory and Esthetic Perception*.

Building on that theoretical framework, our hypothesis is that much of the aesthetic or experiential content of the commercial is carried in the executional peaks. We know from practical experience that these peaks are the moments the consumer actually focuses on the dramatic highpoints of a storyline, or moments of particular emotional or sensory appeal. In basic terms, product peaks contain the rational content of the ad, while executional peaks convey the emotional or entertainment content of the ad. ▶

TABLE 3

Diagnostic relationships to audio-visual measures

	Audio			Visual		
	Early audio category mention	Early audio brand mention	Total audio brand mentions	Total number of peaks	Number of executional peaks	Number of product peaks
Entertaining	-0.35**	-0.39**	-0.06	0.42**	0.54**	-0.18
Interesting	-0.29**	-0.23*	-0.08	0.31**	0.33**	-0.01
Involving	-0.27*	-0.28*	0.90	0.38**	0.44**	-0.13
Different from other ads	-0.20*	-0.16	-0.17	0.21*	0.20*	0.00
Unique	-0.30**	-0.18	-0.04	0.41**	0.43**	-0.02
Boring	0.19	0.29**	0.32**	-0.37*	-0.41**	0.06
Ordinary	0.22*	0.15	0.29**	-0.23*	-0.26*	0.00

Note: * > 95% confidence ** > 99% confidence

Looking at Figure 1 again, we note that there are three peaks in the Flow of Attention® – one product peak and two executional peaks. The product peak contains a brand mnemonic that summarises Degree's positioning – the idea that when your body heat rises, Degree kicks in to work harder. The two executional peaks contain dramatisations of that idea. The first executional peak shows a man running up a flight of stairs to attend a conference. The need for the brand is conveyed in the symbolism of the rising stairs – his body heat is rising as he runs, late for his meeting. The second visual contains dramatic proof that the brand works – he looks cool and collected on the platform as he begins his speech.

Using this classification system, we used two independent coders to perform a content analysis of our sample of 60 commercials. An analysis of the visual content of these commercials provides additional insight into recall and attention scores. Table 3 shows the correlations between the number and type of visual peaks in our sample of commercials and their recall and attention scores.

The total number of peaks is indicative of the visual complexity or information content and narrative structure of a commercial. This measure is not predictive of recall, nor is the number of executional peaks related to recall. But, if we look at the number of peaks devoted to product-related content, we do find a significant correlation with recall. Once again, therefore, we see that day-after recall is related only to explicit

TABLE 4

Correlations between liking and 'breakthrough' measures

	Ipsos ASI recall	Ipsos ASI attention	Ameritest attention	Millward Brown attention
Liking	-0.39*	-0.14	0.38*	0.37*

Note: * > 99% confidence

product-related visuals or 'semantic' information content.

In contrast, the number of product peaks is not correlated with any of the attention measures. Instead, attention is driven by the executional content of the commercial. Commercials that have more peaks in executional content are more likely to capture the attention of the consumer. And, as you can see in Table 3, it is this peak executional content that characterises commercials that are more entertaining, interesting, involving and unique, and less boring and ordinary.

Recall and liking

Finally, if we examine the variable of liking, we find a different result for our sample of commercials; than has previously been reported. There is a significant negative correlation of -0.39 between recall and liking for this sample of 60 commercials. An inspection of the correlations for new product versus established brand or animatic versus film segments of our sample yielded similar negative correlations to recall (Table 4).

Conclusion

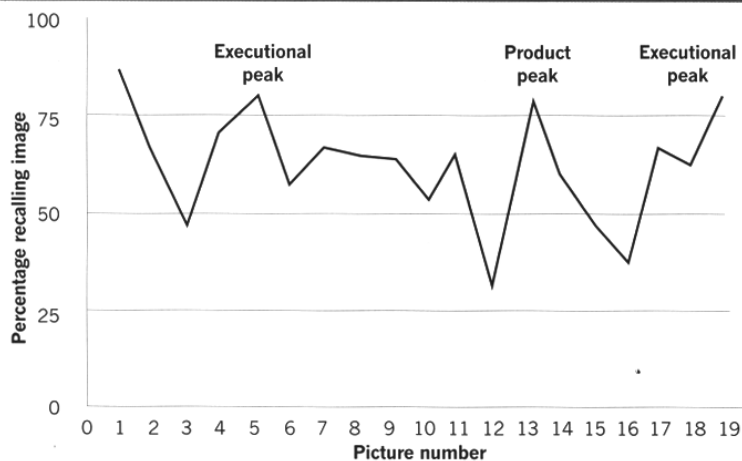
We wanted to explore empirically the meaning of several important advertising

constructs used by advertising researchers. We found that, even though they are both considered measures of breakthrough power, recall and attention measure fundamentally different things. Recall is perhaps a measure of the effectiveness of a commercial when viewed as a sales presentation of product information being made for the consumer. Attention, even when measured in a variety of different ways, is about advertising that is interesting, involving and unique. Attention is characterised by peak executional content, the aesthetics of an ad. Recall scores are completely unrelated to this aspect of television commercials.

Undoubtedly, recall is an appropriate measure for certain advertising applications. However, it appears to reward only one kind of advertising, where the brand is cued early and often and the focus is on communicating rational product 'news' – in short, a linear or vanilla commercial structure. It does not reward other creative formats, for example those involving a reveal-type structure. By rewarding advertising that is boring, ordinary and, indeed, even 'un-likeable', we suggest that the usefulness of recall scores is severely limited. Indeed, the prevailing use of recall testing is certainly one of the greatest sources of frustration for the US advertising creative community. ■

FIGURE 1

Flow of Attention® for Degree deodorant 'meeting'



A Moles: *Information theory and esthetic perception*. Joel Cohen, Trans. U of Ill Press, 1968.

C Young and M Robinson: *video rhythms and recall*. *Journal of Advertising Research*, 29 (3) 1989.

C Young and M Robinson: *'The visual experience of new and established product commercials'* *Advances in Consumer Research*, 18 1991.

H Zielski: *Does day-after recall penalize 'feeling' ads?* *Journal of Advertising Research*, 22 (1) 1982.

john.kastenholz@unilever.com
 chuck@ameritest.net
 Graham.kerr@
 us.millwardbrown.com