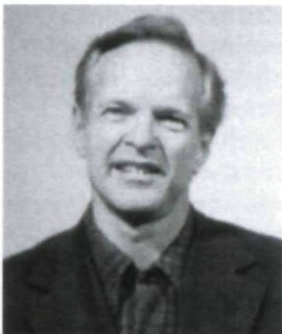


RUSSELL I. HALEY
AND
ALLAN L. BALDINGER

THE ARF COPY RESEARCH VALIDITY PROJECT



RUSSELL I. HALEY is Professor Emeritus of Marketing at the University of New Hampshire. Since 1972 he has also headed a consulting firm, Russ Haley & Associates Inc. He has been a vice president of the D'Arcy and Grey Advertising agencies, and a past president of the Copy Research Council of New York.



ALLAN L. BALDINGER is senior vice president, director of marketing research at the Advertising Research Foundation. He is responsible for overseeing the activities of the ARF's councils and steering committees involved in marketing research. Prior to joining the ARF in 1986, he was the market research director at the Mennen Company for six years.

An earlier version of this article, as presented at the ARF Copy Research Workshop in July 1990, was reviewed by a six-member technical subcommittee of the ARF Copy Research Council, which provided input into that presentation and agreed to its release. This article was reviewed by Ted Dunn, the ARF's technical research consulting director, and three other interested professionals.

The ARF Copy Research Validity Project, which was completed in 1990, had its roots in a speech made by Ted Dunn at the ARF Annual Conference in 1977. Addressing the issue of the validity of copy testing he hypothesized that the answer existed in the archives of copy-testing experience but that, because it took so long to accumulate large enough data bases to permit generalizations and because each copy researcher saw only a small piece of the total, we might never learn the "truth." He called for the formation of a committee to survey advertisers about the amount of validation work that had been done and their willingness to submit their copy-testing files to the ARF—given strong guarantees of confidentiality.

Accordingly, such a committee was formed and, in accordance with time-honored principles, Ted Dunn was named chairman.

A three-year effort devoted to finding available cases bore little fruit. Cases were either unavailable due to confidentiality restrictions, skewed toward certain companies and brands, or unable to discriminate based on sales.

At this point the committee was convinced that the answers everyone was seeking were *not* in the files, at least not in the available files. But it had become even more convinced that the issue of copy-testing validity was an important one, that finding out more about it would be a service to the industry, and that the ARF

was the logical organization to design and oversee research in that area.

The next question was how best to approach something then being called the "Forward Experiment" to differentiate it from any approach that involved the collection of past experience. The committee quickly discovered that a complete experiment covering all aspects of copy research would involve some 6,480 design cells and would cost in the neighborhood of \$2 billion. So a design committee was appointed to develop an affordable compromise.

In June 1982, the design committee came up with preliminary design specifications reflecting various combinations of on-air versus off-air, recruited versus self-selected audiences, in-program versus naked contexts, and single versus multiple exposures. They called for split cable (no matched markets), a minimum test length of six months, and a target of 10 pairs of commercials that were producing significant sales differences. In November 1983, a copy research workshop audience was polled on the kinds of copy-testing contrasts that were of most interest to them. In rank order those were:

- Recall vs. Persuasion
- On-Air vs. Off-Air
- Recruited vs. Self-selected Audiences
- Single vs. Multiple Exposures
- In-program vs. Naked Testing Environments
- Pre/post vs. Post-only Designs

Judging from the quantity of votes the first two of these topics were clearly considered the most important.

A funding drive was then launched involving a six-cell design (three off-air and three on-air) and a financial target of approximately \$750,000. However, by November 1984, it was apparent that the drive would fall short and that design changes would be required if the project were to be brought to fruition. Fortunately, commercial copy-testing firms stepped in and assumed the responsibility for conducting, at their own expense, the on-air tests. They were assured that their identities would be carefully protected regardless of the outcome of the testing and, in turn, agreed not to promote the results of the tests no matter how well their service fared.

The design committee also developed a cost-efficient alternative. Rather than simply putting pairs of commercials that had been copy tested into very expensive sales tests, it recommended that the process be reversed. In other words, the sales-test results would be examined first, and only those pairs that were producing significantly different sales results would be copy tested. This avoided situations in which copy tests showed differences but sales tests did not. It was felt that, while that kind of situation is more likely to occur, it offers less opportunity for learning. After considerable discussion, it was also decided that to expand the design to include this circumstance (i.e., where no sales difference occurred) would be of less interest to the community and be beyond the financial resources available for the study. So the objective of the final design became "to see how well the types of copy tests in common use can identify

known sales winners from pairs of commercials for specified brands."

The committee then began to look for pairs of commercials that were producing significant sales differences with a target of 10 such pairs. Once an eligible pair was located, those commercials were quickly copy tested in markets that were close geographically to the markets from which the sales information had come but at the same time far enough away so that the people in those markets had not been exposed to the commercials.

Split-cable research companies, of course, could not divulge the names of client companies that were running copy tests so the committee contacted clients directly. For a year and a half the research directors of companies thought to be using split-cable copy testing were regularly contacted to determine whether or not they had an eligible pair of commercials and, if so, whether they would be willing to release that pair to the ARF for its own copy testing.

As mentioned above, the goal had been to find 10 eligible pairs. However, it was quickly found that the conditions placed on eligibility made pairs of commercials hard to locate. To be eligible the pairs had to be commercials not previously aired, from an advertiser making minimal use of print in the test markets, commercials that were the only ones in use during the test, and ones for which at least six months' (and preferably a year's) sales data were available. While it was apparent that strict adherence to these criteria would substantially lengthen the period needed to obtain pairs, it was decided that the ideal of relatively clean tests was worth the wait.

By 1989 five pairs had been obtained and tested. Some had been lost to fast rollouts and

some were rejected because the sales differences, while statistically significant, were only one or two percentage points. It was getting progressively more difficult to locate eligible pairs and funds for testing purposes were running low. So it was decided that results would be reported for the available cases with the proviso that, if additional cases turned up and industry interest levels were high enough to indicate that additional funding would be available, the project might be extended.

A top-line report on findings was made at the ARF Annual Conference in April 1990 and a final report at the ARF Copy Research Workshop in July 1990.

Objectives and Research Designs

The objectives of the research that emerged from this process can be summarized as follows:

- How well do copy tests, as presently conducted, identify known sales "winners"?
- Which individual measures do the best job?
- Which general *types* of measures are most predictive?
- Are on-air designs preferable to off-air?
- Are pre/post designs preferable to post-only (matched group) designs?
- Are multiple-exposure designs better than single-exposure designs?
- Is any one copy-testing *system* superior to the others?

The final research design involved six copy-testing methods, three of which were off-air and three on-air. Ten commercials (five brand pairs) were tested via each of the six copy-testing methods; for each pair it was known that one of the commercials was producing significantly

more sales than the other (by margins from +8 percent to +41 percent). So the study design consisted of 30 cells and looked as shown in Figure 1.

Between 400 and 500 people were interviewed in each cell, thus totaling between 12,000 and 15,000 interviews. All respondents were members of the target groups as defined by the brands being advertised. In most cases these were simply category users.

The products advertised were all packaged goods in the food and health-and-beauty-aids categories. The brand names were all established brand names, and for each of them television was the predominant form of advertising.

The copy-testing measures that were built into the questionnaires for the off-air cells fell into six general types: measures of persuasion, brand salience, recall, communications (playback), overall commercial reaction (liking), and commercial diagnostics.

Within each general type were a number of specific measures as follows:

Individual Measures

- (1) *Persuasion*
 - Brand Choice
 - Constant Sum
 - Purchase Interest (top box)
 - Consideration Frame
 - Overall Brand Rating (top box and average)
- (2) *Salience*
 - Top-of-mind Awareness
 - Unaided Awareness
 - Total Awareness (Unaided Plus Aided)
- (3) *Recall*
 - Recall Brand from Product Category Cue
 - Recall from Brand Cue
 - Claimed Recall from Full Set of Cues
 - Related Recall from Product Category Cue
 - Related Recall from Full Set of Cues

Figure 1
Copy Testing Method

	Off Air			On Air		
	1	2	3	4	5	6
Pair A						
Pair B						
Pair C						
Pair D						
Pair E						

- (4) *Communication*
 - Main Point Communication
 - Ad Situation/Visual
 - Ad Positive
 - Sales Point Communication
 - Main Point Situation
- (5) *Commercial Reaction (liking)*
 - One of the Best Seen Lately (top box and average)
 - Like/dislike (top box and average)
- (6) *Commercial Reaction (diagnostics)*

Positive

 - I learned a lot from this advertising.
 - Tells me a lot about how product works.
 - Helps me to find the product that I want.
 - Told me something new about the product that I didn't know before.
 - This advertising is funny or clever.
 - I find this advertising artistic.
 - This advertising is enjoyable.

Negative (Items reversed)

 - The product doesn't perform as well as this ad claims.
 - This ad doesn't give any facts, just creates an image.
 - This advertising insults the intelligence of the average consumer.

- This advertising is boring.
- This advertising is in poor taste.

Detailed descriptions of the on-air cells are not available for reasons of maintaining confidentiality. The research firms involved used their normal procedures. The three off-air systems were a pre/post system, a post-only system, and a multiple-exposure (re-exposure) system, as shown in Figure 2. Respondents for the off-air cells were recruited in shopping centers and exposed individually to a 10-minute program that contained a single three-commercial pod, one of which was the commercial of interest with the remaining two included to provide noncompetitive clutter. All commercials were 30 seconds in length. Respondents were recruited to watch the program (not the commercials). Interviewing took about 30 minutes in total. During the questioning, two masking product categories were used to disguise the product category of interest. A 100 percent validation was done on the interviews.

All studies have limitations and the Copy Research Validity Project is no exception. In interpreting the results a number of

things must be kept in mind. First, all the ads tested were for packaged goods. There were none for services, hard goods, or retailers. Second, no new products were included. Third, the tested commercials were chosen by availability. No attempt was made to stratify them in any way. (Nevertheless, the executions were quite varied. They included hard sell and soft sell, voice-overs and on-camera spokespersons, music and no music, outdoor and indoor settings, humorous and serious commercials, males and females, and adults, teens, and children.) Fourth, as noted earlier, pairs of commercials which showed insignificant sales differences were avoided. Fifth, the copy tests followed the obtaining of sales measures but were done in markets with no past exposure to the copy. Sixth, the sample size of five pairs is smaller than we would have liked (although even with five pairs the chances of picking five winners out of five by luck alone are only one in thirty-two). And, finally, the large number of alternatives tested increases chances of spurious significance for some measures.

On the other hand, the study has a number of obvious strengths. It is the only public study in which fresh copy executions of individual commercials for established brands have been evaluated both by extensive copy testing and by split-cable sales (with post adjustment for in-store factors). This, of course, allows individual copy-testing measures to be directly compared to sales performance. Thus it relates directly to the question, frequently faced by advertisers, of "how valid is copy testing when it is used to select, from two or more copy executions, the one that is likely to have the greatest impact on sales?"

The use of pairs of commercials for the same brands effectively removes potentially large brand effects from the study.

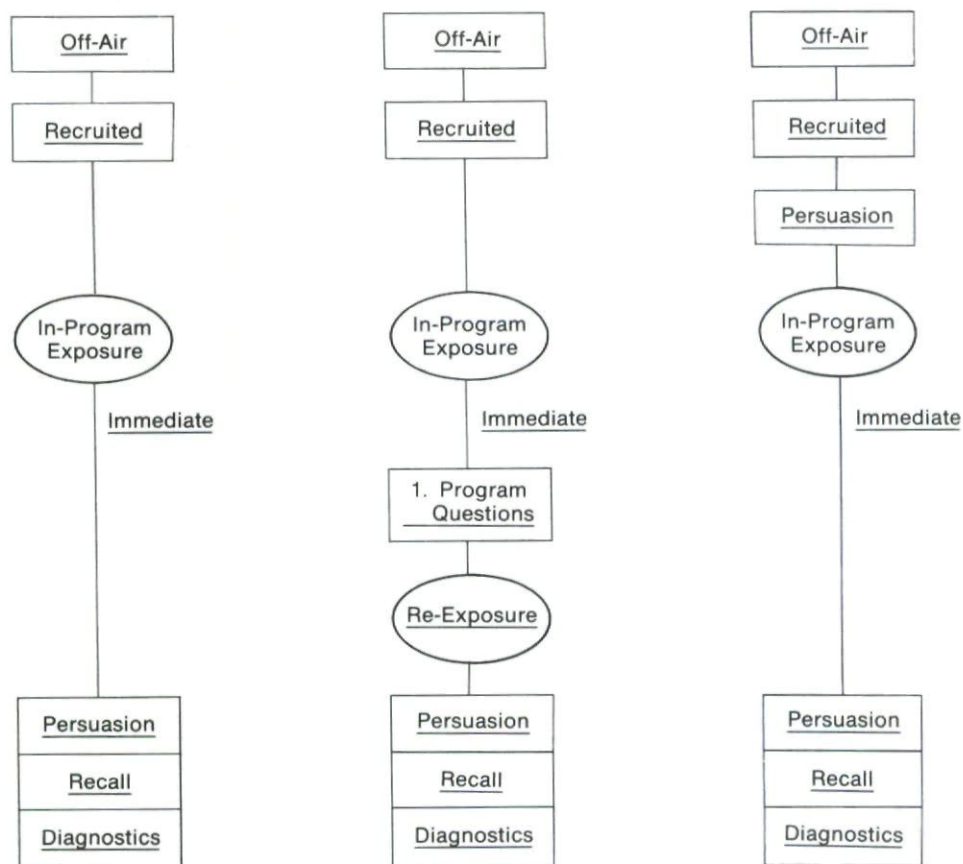
All major types of commonly used copy-testing measures were represented in the research, although not necessarily all variations within types. Copy-testing firms can and do claim proprietary differences. Large and small brands were also represented (market shares ranged from 7 percent to 35 percent). And a blue-ribbon interindustry steering committee oversaw both the design and execution of the research. The Marketing Science Institute also reviewed the design, approved it, and recommended that their members support the project financially.

Principal Findings: Measures

The data presented in this section necessarily involve some arbitrary decisions. One concerns the length of time during which sales can be expected to be influenced by the differences in copy that are reflected by the copy-testing measures. A one-year time period (or as close to it as possible) has been chosen for purposes of this review. Had six months been chosen results might have been slightly, although probably not materially, different.

Even more difficult choices exist with respect to handling questions of statistical significance. The samples for each

Figure 2
ARF Copy Research Validity Project



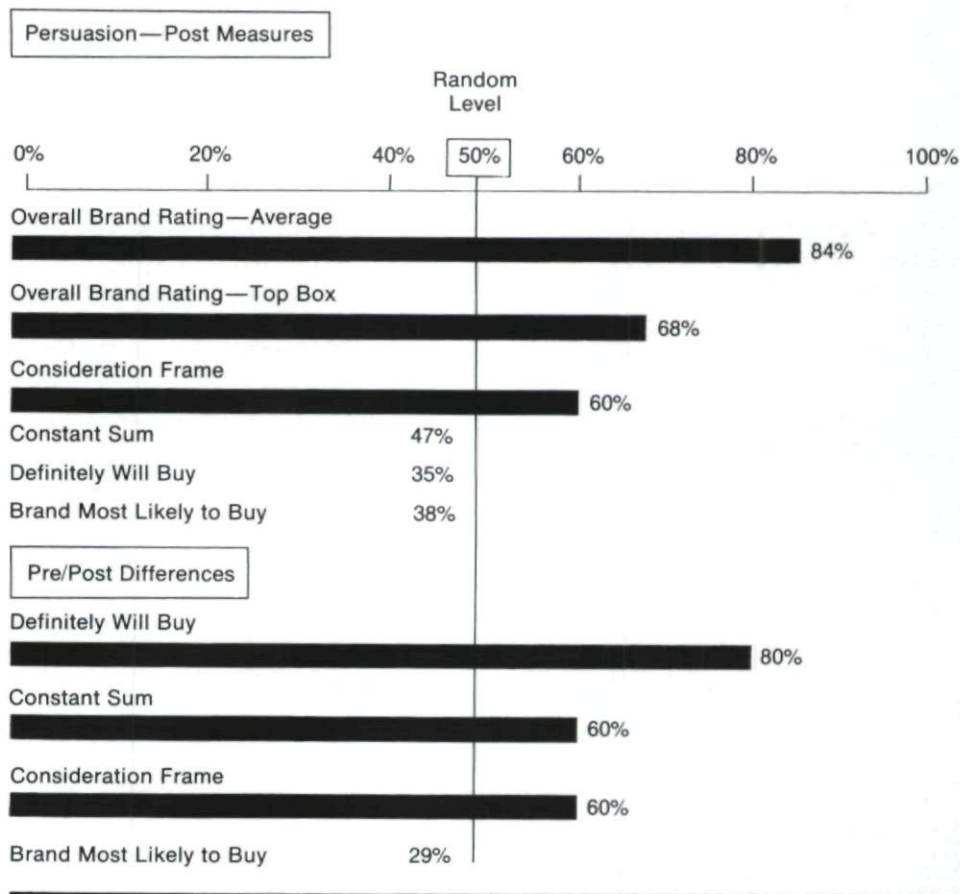
commercial test are independent, and therefore it is legitimate to aggregate results for individual measures. However, there are two sources of intercorrelation that affect other types of aggregation. Measures of the same general type (e.g., purchase intent and overall rating) are correlated and therefore should not be added together as though they were independent. Likewise, tests in the six copy-testing systems were all done with the same 10 commercials. Again this results in some level of intercorrelation and argues against aggregation.

However, there are also advantages to the design chosen. Testing the same 10 commercials in six different methods provides the opportunity of replicating the predictive values of the measures studied. And, it is also desirable to present the results of studies as complex as this one in as simple and straightforward a manner as possible. So, given the fact that individual readers are likely to be interested in quite different levels of aggregation and to favor different levels of statistical significance, we have opted for the sacrifice of some statistical rigor in the interest of clarity. The net result is to increase the standard errors shown for aggregated measures. A separate table shows average levels of intercorrelations so that rough judgments can be made of their effects. (See the Appendix.)

There are two somewhat different ways of evaluating the results of comparative copy tests where the question is whether copy A is better than copy B. The scientific approach is to choose an acceptable level of statistical significance, say the 80 percent level, in advance of the test. Then, if one piece of copy surpasses the other at that level, it is considered the winner. And if neither piece of copy is superior

Figure 3

Persuasion Percent Directional Wins: 50% Confidence Level



to the other, the results are thrown out and the choice between them is made on the basis of judgment.

A perhaps more pragmatic approach, given belief in the copy-testing measure employed, is to choose a criterion measure or measures and to consider that whichever piece of copy scores higher on that measure is the winner—even if the margin of superiority is a single point. The rationale for this approach is that, if the copy-testing measure is valid, the higher scoring execution always has a higher, if only slightly higher, probability of being the better commercial.

Also, there are some political reasons why this approach is sometimes favored. If a brand manager should decide to run

the lower scoring commercial and if sales of the brand subsequently decline, he or she is vulnerable to second-guessing from top management. If, on the other hand, the higher scoring commercial is chosen the research can always be used to support the soundness of the brand manager's decision. So higher scoring commercials are usually the ones put on the air even if their margin of superiority is a narrow one.

Results are presented for both the scientific and the pragmatic approaches. In the interest of succinctness they have been combined into twelve figures, two for each of the six areas of coverage. The first set of two figures deals with persuasion measures (see Figures 3 and 3A). Figure 3 is appropriate for people

who lean to the pragmatic approach. It is simply that percentage of time that the measure was able to correctly identify the better selling commercial of its pair. In other words, the "average overall brand rating" was able to identify the better selling commercial in 84 percent of the pairings in which it was used.

Figure 3A pertains to the scientific approach, using the 80 percent level of significance. The data shows that 37 percent of the time the differences in the Overall Average Brand Ratings of the pairs were not only in the right direction but were large enough to be statistically significant at the 80 percent level.

The random level of correct predictions at the 80 percent level is, of course, 20 percent. To make the results easier to understand, they have been converted to indexes in the first right-hand

column. Thus, the 37 percent, when divided by the random level of 20 percent, becomes an index of 184.

The next column in Figure 3A shows the number of methods, out of six, that contained this particular measure. Since four measures out of six contained the overall brand rating, and there were five pairs of commercials, this measure was included in about 20 (i.e., 4×5) of the total cells.

The final column in Figure 3A contains T-values that allow readers to choose any level of significance preferred. For example, values that exceed 1.28 are significant at the 80 percent level for a two-tailed test of whether the observed level of correct prediction is higher than the random level of 20 percent; significant differences at this level are shown in index form and are boxed. T-scores are shown for

those interested in performing statistical tests at levels other than 80 percent. For example, a difference at the 95 percent confidence level (two-tailed) requires a T-value of 1.96.

The Overall Brand Rating question was asked as follows:

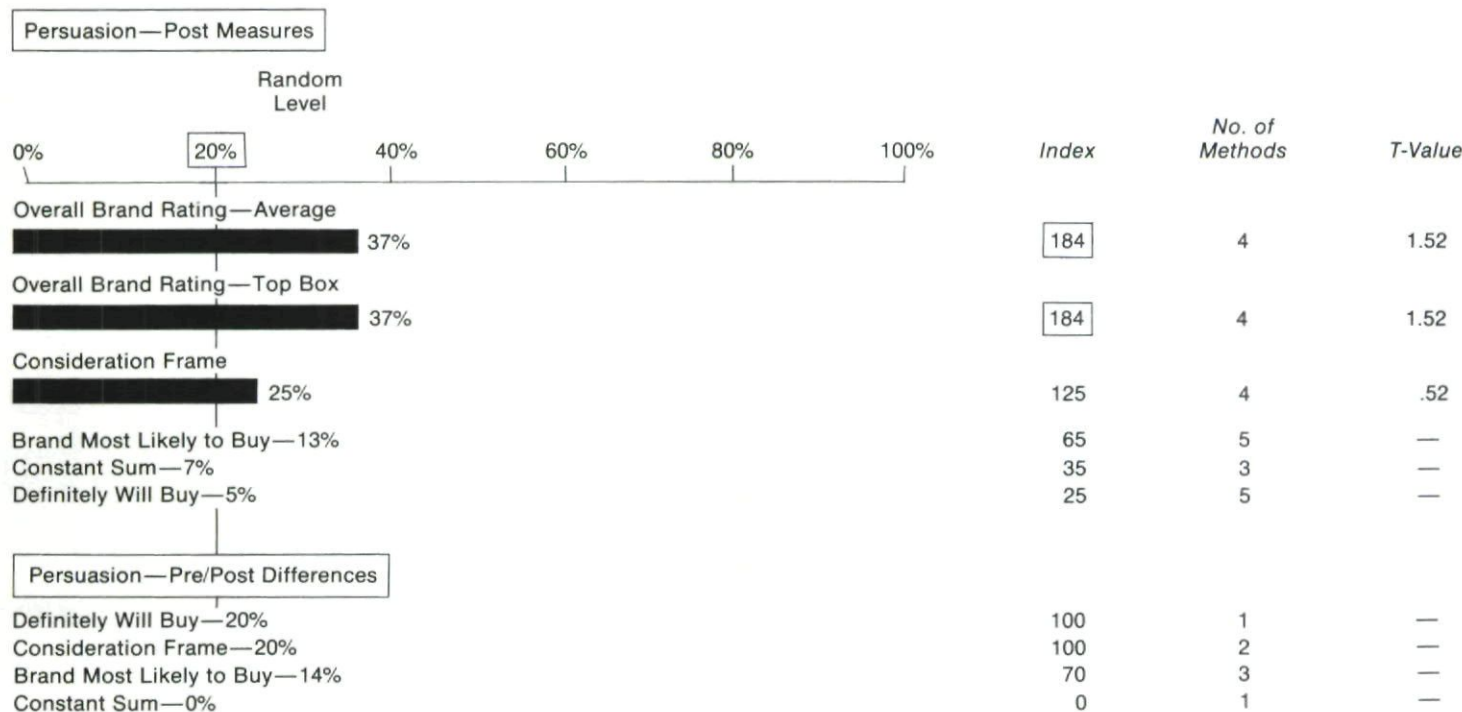
Overall (Average)

"Based on the commercial you just saw, how would you rate the brand in the commercial on an overall basis using the phrases on this card?"

- Excellent 6
- Very good 5
- Good 4
- Fair 3
- Not so good 2
- Poor 1

In the area of Salience, the strongest performance came from Top-of-mind Awareness (see Figures 4 and 4A). It correctly identified the winning commer-

Figure 3A
Persuasion Percent Significant Wins: 80% Confidence Level



cial 73 percent of the time and had an index of 167, but was not significant at the 80 percent level of confidence. It was asked in the following way:

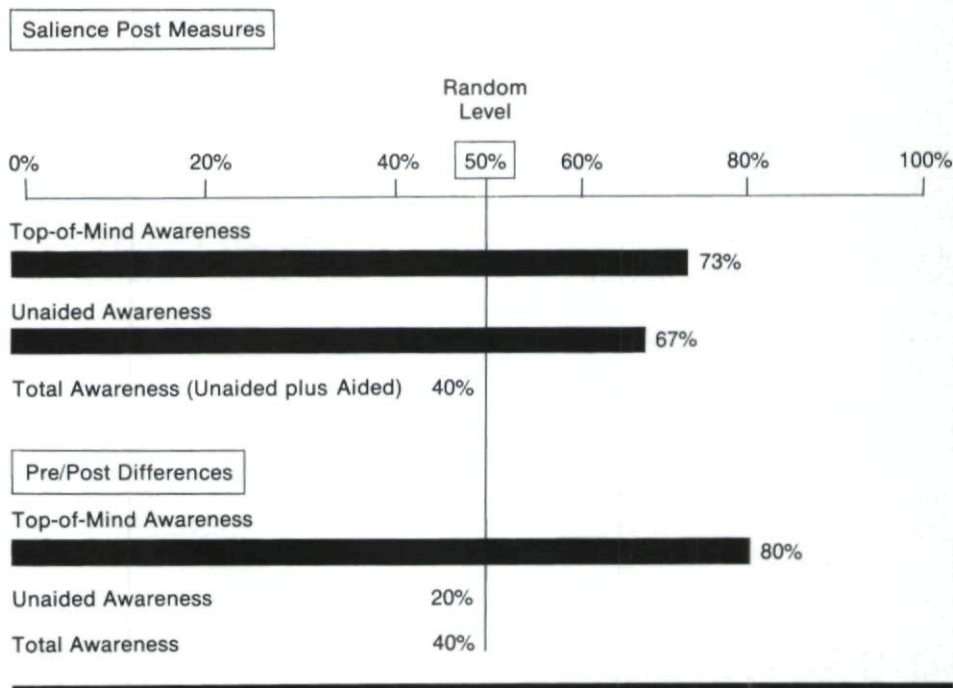
Top-of-Mind Awareness

“When you think of (Product Category) what are all the brands you think of?”

First Mention _____

The next set of measures calculate recall in a variety of ways (see Figures 5 and 5A). The measures that perform best use minimal cues. It appears that extensive probing to enlarge the recall base is not a good idea. While probing does provide more people who can be classified as recallers and who can be contrasted with nonrecallers for insights into the sources and effects of recall, it does so at the risk of diluting the predictive power of the recall measure. The measure that performed best in this set was recalling the brand with only a product category prompt. It di-

Figure 4
Saliency Percent Directional Wins: 50% Confidence Level

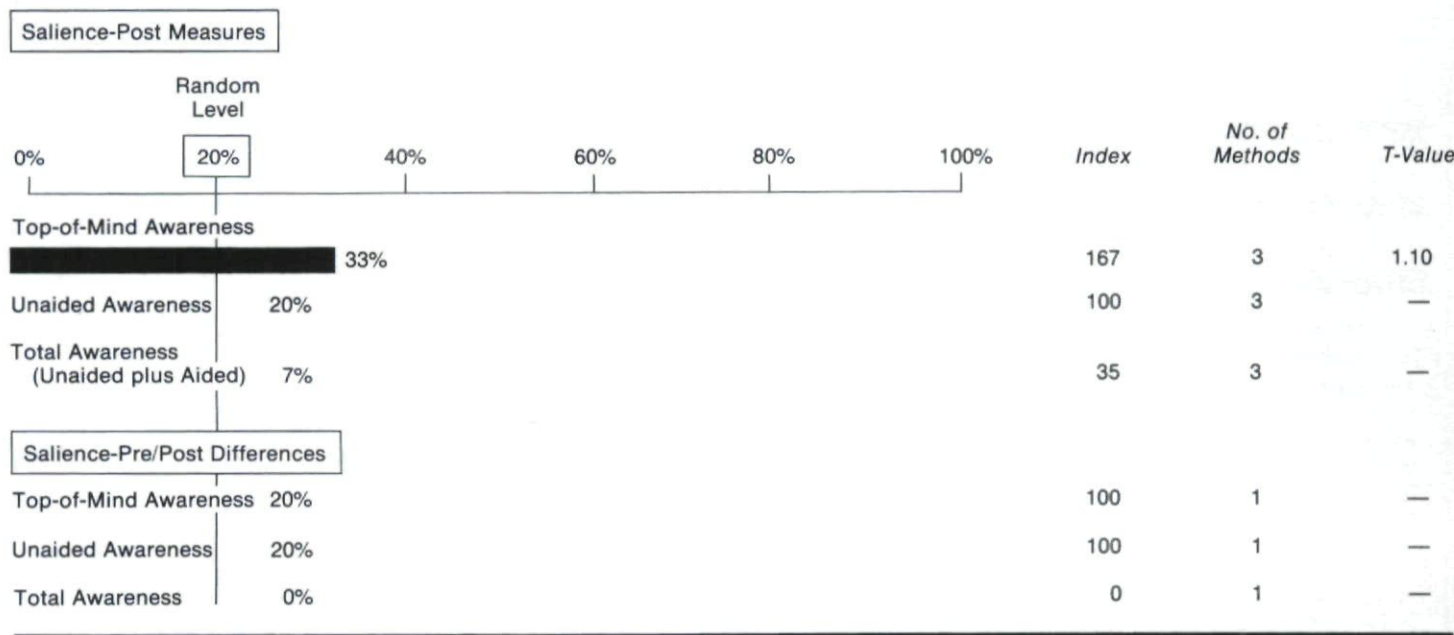


rectionally picked the winning commercial 87 percent of the time and had an index of 234. It was asked as follows:

Recall Brand from Product Category Cue

“On the show you just saw, do

Figure 4A
Saliency Percent Significant Wins: 80% Confidence Level



you remember seeing a commercial for a brand of _____?" (If yes) "What brand was that?" _____

In the area of communications, the best predictor was the Main Point of the Communication (see Figures 6 and 6A). It identified winners 60 percent of the time and had an index of 200. Even better at the pragmatic level was

playback of the ad situation/visuals. It had correct directional predictions 67 percent of the time but a slightly lower index at 188. Both of these measures were gathered via the following question.

Main Point Message

"Of course the purpose of the commercial was to get you to buy the product. Other than that

what was the main point of the commercial?" _____

Next we come to a set of measures that had a better prediction record than any other type of measure included in this research: measures of the overall affective reaction to the commercial (see Figures 7 and 7A). The average impression of the commercial, derived from a five-point liking scale, picked sales winners directionally 87 percent of the time and had an index of 300 (i.e., picked winners and was significant 60 percent of the time). An agree/disagree scale for the statement "This ad is one of the best I've seen recently" picked even more winners (93 percent), but its index of significant wins was a little less impressive (200). This is how the overall affective reaction to the commercial was measured.

Impression of Commercial (Top box)

"Thinking about the commercial you just saw, please tell me which of the statements on this

Figure 5

Recall Percent Directional Wins: 50% Confidence Level

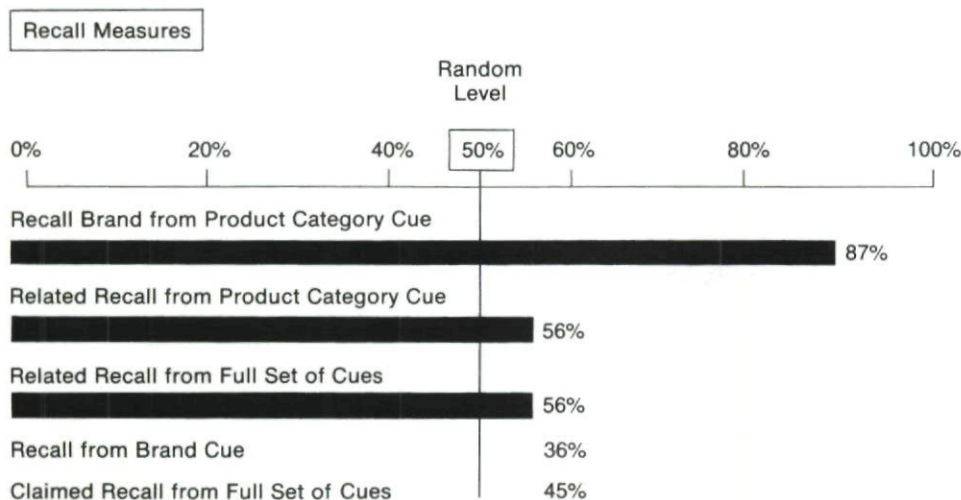


Figure 5A

Recall Percent Significant Wins: 80% Confidence Level

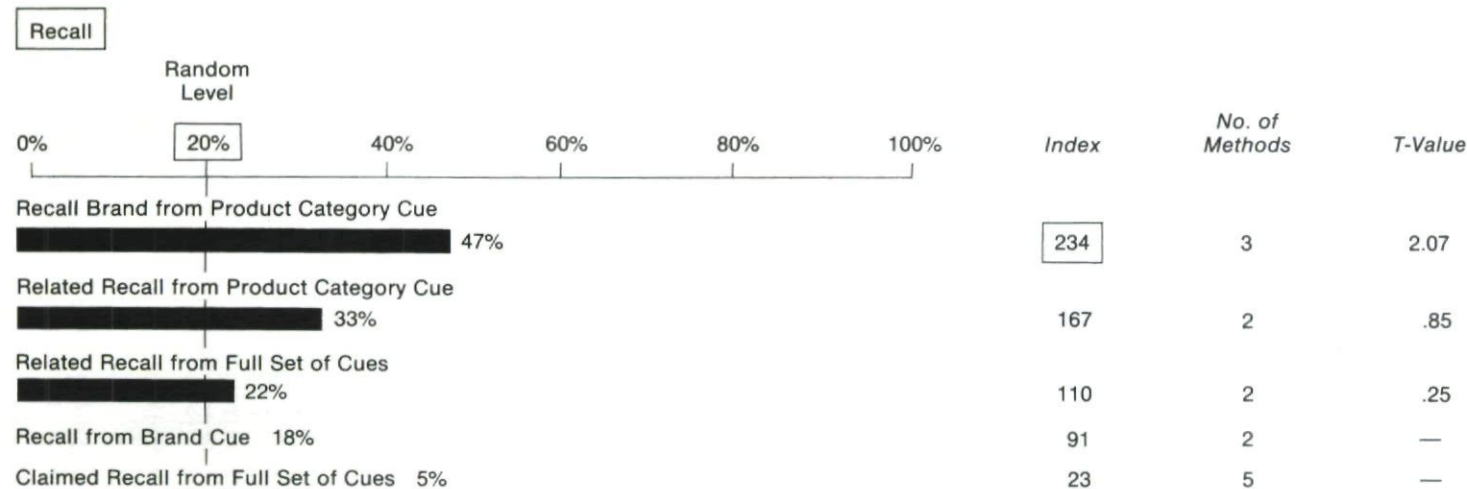
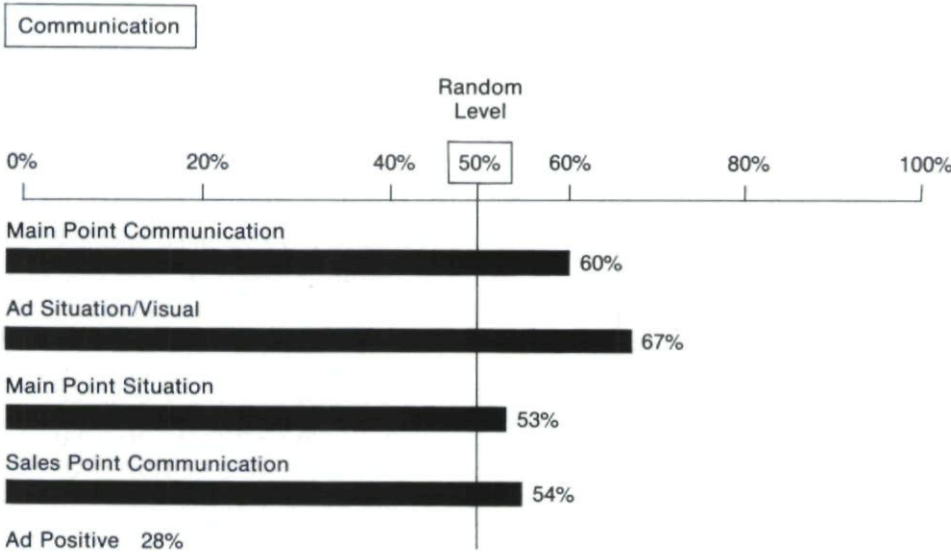


Figure 6

Communication Percent Directional Wins: 50% Confidence Level



card best describes your feelings about the commercial."

- I liked it very much 5
- I liked it 4
- I neither liked it nor disliked it 3
- I disliked it 2
- I disliked it very much 1

The last group of measures are

ones that historically have been called "diagnostics." These were thought to be measures that were not in themselves evaluative but that could explain why other more purely evaluative measures behaved as they did. Clearly some rethinking on that score is needed. Many of the "diagnostics" showed strong relations

to sales. (See Figures 8 and 8A.)

Among positive diagnostics the "news" item ("Told me something new about the product that I didn't know before") identified winners 87 percent of the time and had an index of 200. And among negative diagnostics "This ad doesn't give any facts, just creates an image" picked losers 60 percent of the time and had a significance index of 234. (See Figures 9 and 9A.)

The study from which these items were drawn is entitled "Advertising and Consumers" and was coauthored by Rena Bartos and Ted Dunn and was published by the American Association of Advertising Agencies in 1976.

Responses to these items were gathered on the following agree/disagree scale:

Scale for Diagnostics

"I am going to read a number of statements which people have used to describe their opinions and feelings about advertising. For the commercial you saw, I would like to know how much each statement describes how

Figure 6A

Communication Percent Significant Wins: 80% Confidence Level

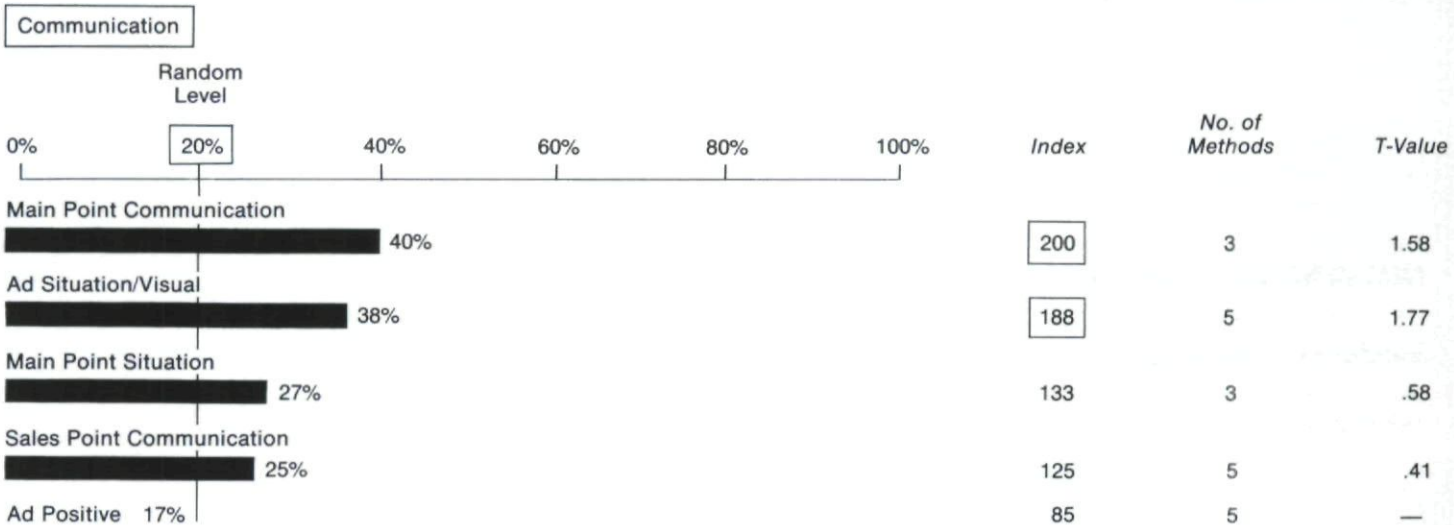
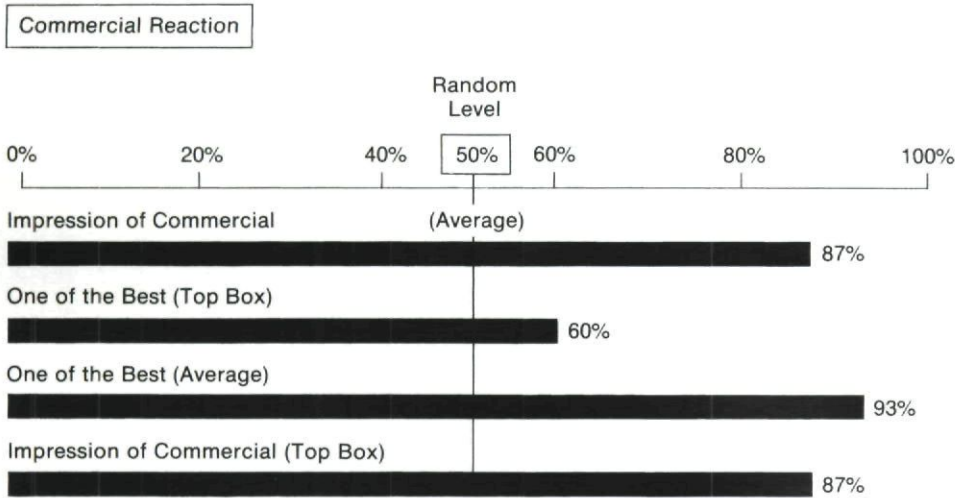


Figure 7

Commercial Reaction Percent Directional Wins: 50% Confidence Level



you feel.”

- Agree completely 4
- Agree somewhat 3
- Disagree somewhat 2
- Disagree completely 1

Summing up this section, it is apparent that copy-testing measures can predict sales, when sales differences are present (i.e., copy testing works). Moreover, there is at least one reasonably successful predictor within each

of the types of measures covered. So all types of measures currently being employed are potentially useful.

The Possibility of Interaction between Methods and Measures. It is quite likely that the measures that rank highest in Figures 3 and 3A through Figures 9 and 9A are those that are most robust to the kinds of copy-testing environments in which they are used. That leaves open

the question of whether some individual measures do particularly well in some copy-testing methods.

To shed some light on that question, Figure 10 shows every measure that had perfect (five out of five) directional predictions in at least one of the three off-air methods. In order that these measures can be compared across methods, their results for all three off-air methods are shown. Boxes with “x’s” in them indicate correct predictions and those that are blank incorrect predictions.

Thus, within the pre/post off-air methods the overall average brand rating was perfectly predictive in all five pairs. Likewise, it had a perfect record in the post-only design. But in the re-exposure design it was directionally correct for only four of the five product pairs.

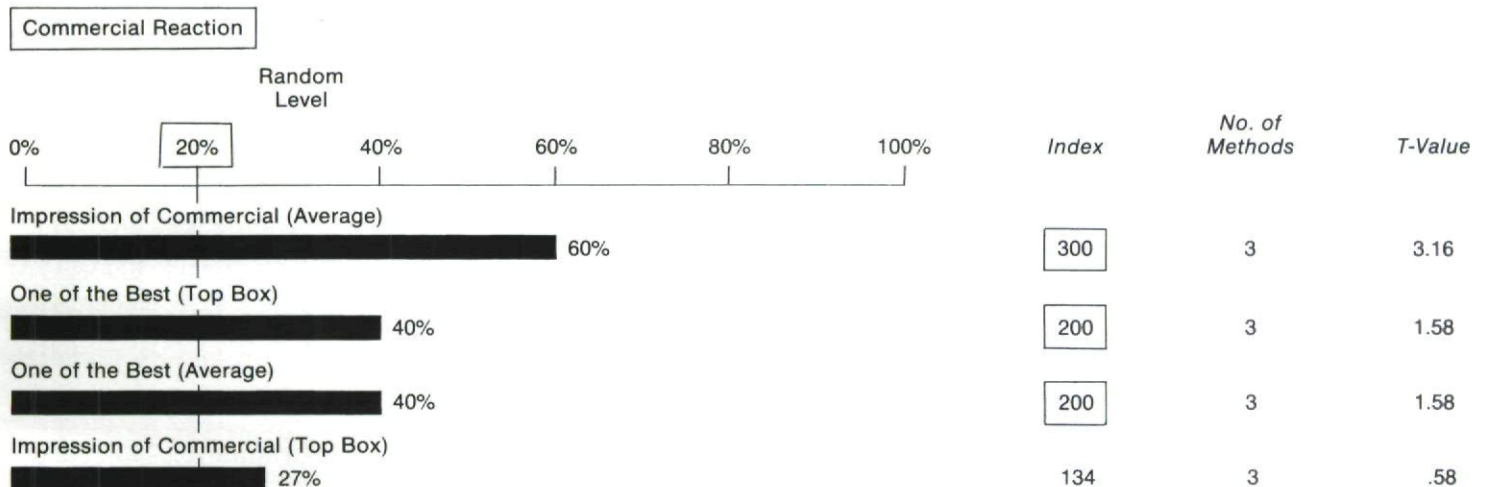
No consistent pattern emerges. However, most of the measures that did well in one method appear to have done well in the others also.

Two-variable Predictors

Because some individual measures show strong performance,

Figure 7A

Commercial Reaction Percent Significant Wins: 80% Confidence Level



the question naturally arises whether some two-variable predictor might not do a perfect job. Reinforcing the value of this question is the fact that some large marketers and testing services multiply their persuasion scores times their recall score to obtain an overall score reflecting their composite effects.

To investigate the performance of two-variable combinations, the strongest measure from each of the six types of measures, discussed above, was multiplied by the strongest measure in the other five categories.

Looked at in this way, no two measures when multiplied together produce a score that is a perfect predictor for all five product categories in all three of the off-air methods. However, using data at the 50 percent level, five two-way combinations do correctly call 14 of the 15 cases (3 methods by 5 product pairs). The most successful combinations were:*

- Persuasion × Liking
- Persuasion × Not Boring
- Recall × Main Point Communication
- Recall × Liking, and
- Persuasion × Recall

As noted above, the last of these pairs is already being used. It appears probable that "liking" is what Gordon Brown has called a "creative magnifier" for both persuasive messages and for messages that are recalled.

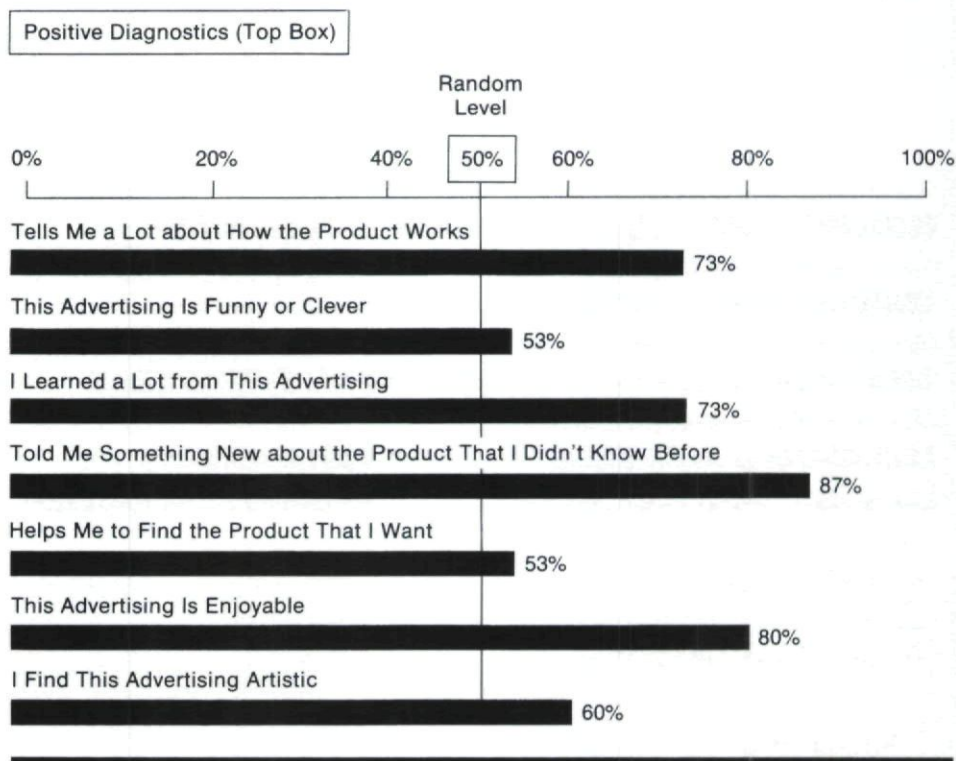
Multivariate Predictors

Another possible approach to understanding the combinations of items that do the best job of explaining the differences between winners and losers is dis-

*Persuasion = Overall Brand Rating; Liking = Impression of Commercial (Average); Recall = Recall Brand from Product Category Cue.

Figure 8

Positive Diagnostics (Top Box) Percent Directional Wins: 50% Confidence Level



criminant function analysis. Under this approach the winning commercials as a group are contrasted with the losing commercials as a group and equations are developed to assign each commercial to the winning or to the losing group using all of the measures included in the study as predictors. This process, however, sacrifices the advantage of direct pair-wise comparisons and assumes instead that the winning commercials as a group share some underlying characterizations that differentiate them from the losing group. The first six variables to enter the discriminant function equation and the cumulative percentages of correct classifications are shown in Table 1.

Factor Analysis

Still another technique useful

for obtaining insights into the nature of the measures included in the study is factor analysis. Several different forms of factor analysis were tried. Two are summarized in this article.

The first included only measures that were common to all of the off-air tests. Four factors accounted for more than 80 percent of the variance in that set. (See Table 2.)

The first factor was named "Informational/Consideration Frame." It links information and the "Overall Brand Rating" persuasion measure. Note also the implication that advertising that does not provide viewers with useful information runs the risk of being considered an insult to their intelligence.

The second factor is called "Positive Commercial Reaction." It is clearly "liking," including the two strong affective pre-

dictors along with the "artistic" and "enjoyable" descriptors. The third factor appears to be

"Persuasion" and links it to entertaining advertising. And, finally, "Brand Recall"

appeared as a stand-alone item. So, summing up, the four underlying factors relate to informa-

Figure 8A

Positive Diagnostics (Top Box) Percent Significant Wins: 80% Confidence Level

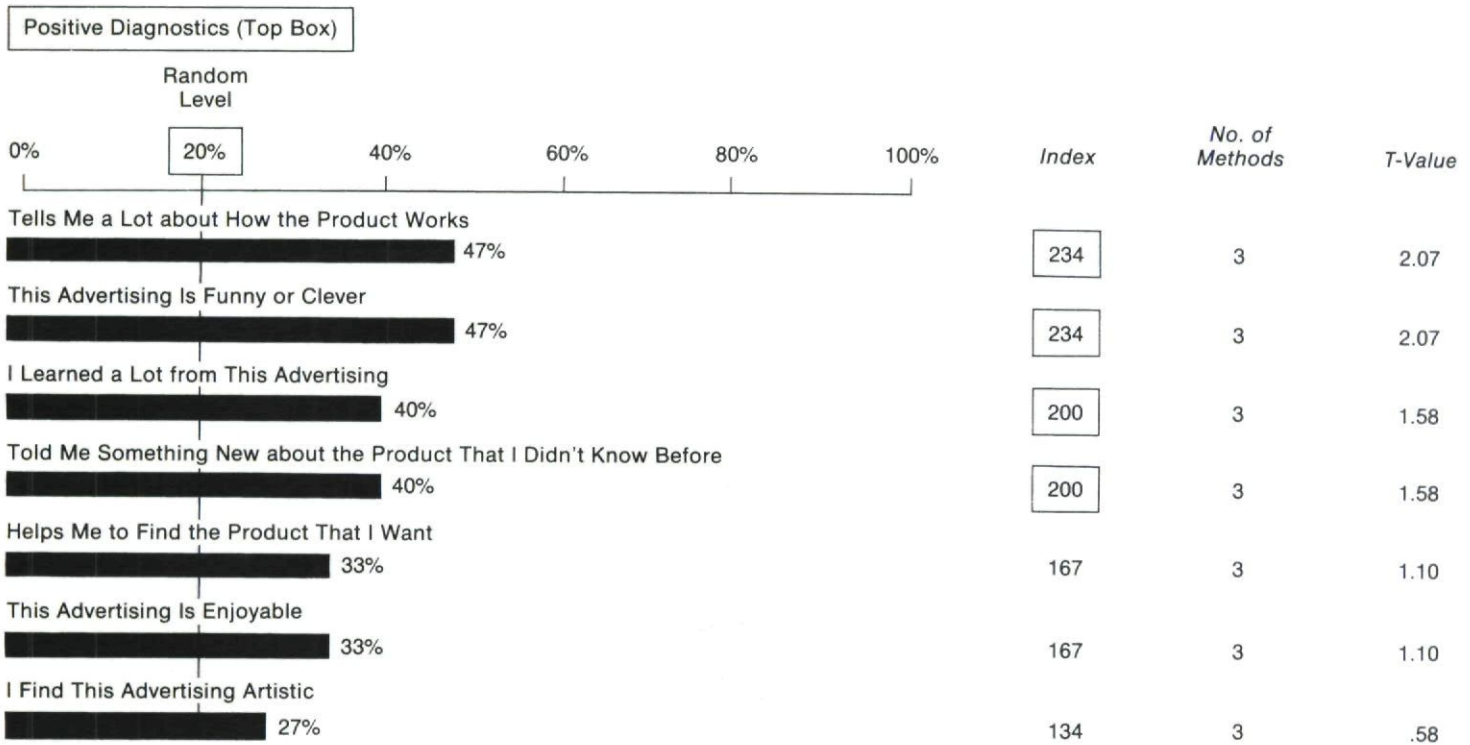
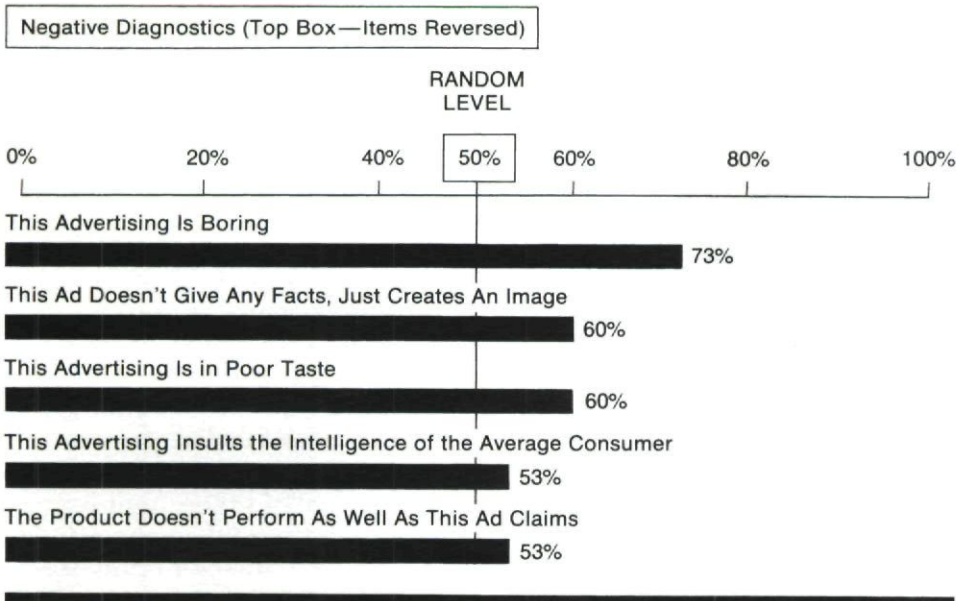


Figure 9

Negative Diagnostics Percent Directional Wins: 50% Confidence Level



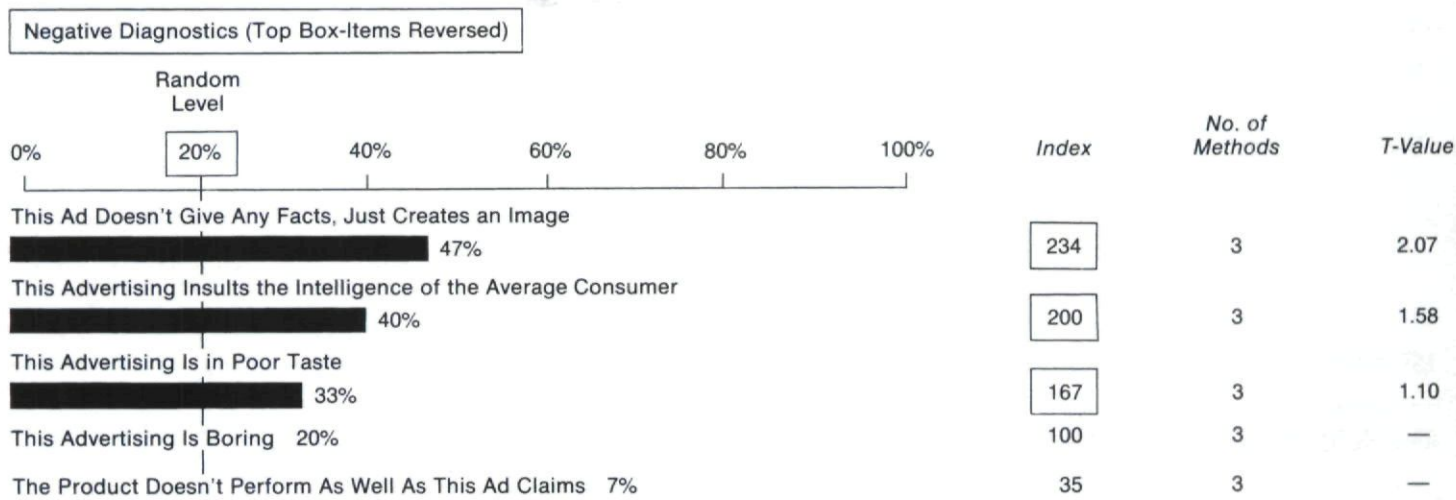
tion, liking, persuasion, and brand name recall. Doubtless their importance varies both with the objectives of the commercial and with the type and status of the brand to be advertised. However, all are potentially important components of advertising and all should be covered in any complete copy-testing method.

Because some measures, notably pre/post choice measures, were included in only one off-air method (the pre/post method), data from that method were factored separately to see what differences might occur. This time two factors were found that, together, accounted for 61.1 percent of the total variance.

The first factor bears a strong resemblance to the first factor in the previous factor analysis and

Figure 9A

Negative Diagnostics Percent Significant Wins: 80% Confidence Level



therefore is given the same name, "Informational/Consideration Frame." Note that the last two measures shown in the set with positive factor loadings are pre/post measures (change in consideration frame and change in Top-of-mind Awareness).

The second factor links persuasion and liking measures and accordingly is called "Persuasion/Liking." Most of the persuasion measures fall out together in this factor. The fact that the Overall

Brand Rating and the Pre/Post Constant Sum appear in the same factor reinforces the conclusion that these are alternative measures of persuasion and in this respect mirrors the findings of the old Arrowhead #9 Study.

Also of interest is the association between persuasion and liking that is reflected in the items included in this factor. (See Table 3.)

Summing up, the various factor analytic approaches sug-

gest that information and entertainment are both consistent underlying components of advertising and that liking, persuasion, and recall are sometimes linked to one, sometimes to the other, and sometimes to both. Because advertising can work in a variety of ways all are worthwhile areas of coverage for any standardized method of copy testing.

Principal Findings: Methods

As mentioned earlier, among the main objectives of this research were findings concerning choices between on-air and off-air designs, pre/post and post-only designs, and single- versus multiple-exposure designs.

The argument over whether it is better to use an on-air or an off-air test is one of long standing. On-air advocates focus on the advantages of "natural" exposure. For these people the gold standard is that people should be exposed to the test commercials in their own homes over their own television sets and that they should not know that they are in an experiment at the

Figure 10

Correct Directional Predictions (By measure,* within design type)

Measurement	Design Type		
	Prepost	Post-Only	Re-exposure
Overall average brand rating	X X X X X	X X X X X	X X X X
Recall brand with Product Category Cue	X X X X	X X X X X	X X X X
Liking, top box	X X X X X	X X X X X	X X X
Liking, average	X X X X	X X X X X	X X X X
One of best, average	X X X X	X X X X X	X X X X X
Something new, top box	X X X X	X X X X X	X X X X
Boring, average	X X X X	X X X X	X X X X X
Insults intelligence, top box	X	X X	X X X X X

* Measures included were those having perfect directional predictions in at least one design type.

time of exposure. It is recognized, of course, that some compromises may have to be made for economic or practical reasons. However, the fewer of these the better.

People favoring off-air tests emphasize the value of being able to control the many variables that can affect the results of testing. For them, the cardinal rule of copy testing is to be sure that the effects you are measuring and any differences that you find are attributable with great certainty to the commercials being tested. They also place emphasis on the flexibility and potential diagnostic power of controlled testing.

Figure 11 shows consolidated results for the three on-air cells in relation to the three off-air cells.

Superficially, results favor the off-air proponents in almost every respect. Measures used in an off-air environment do a better job of picking known sales winners than comparable measures gathered on air. However, there are complications.

It will be recalled that for copy testing done off-air the markets chosen for copy tests were carefully matched geographically with those from which the sales data were derived. And it is a well-known fact that copy-test results for the same piece of copy can and often do vary from market to market. For the off-air methods, however, geographical differences were largely neutralized.

On-air testing, on the other hand, was done as it is normally done—a typical sample being “four geographically dispersed” markets. So, for the on-air tests, the locations of the split-cable markets supplying the sales data were ignored.

Thus, while it is true that the off-air copy tests predicted split-cable market sales best, it could be argued that sales in split-cable

Table 1
Multiple Variables (Discriminant Function Analysis, Individual Variables) All Off-Air Cells

Variables	% Classified correctly
Recall brand from Product Category Cue	73.3
Doesn't give any facts, just creates an image	86.7
This advertising is boring	83.3
I learned a lot from this advertising	86.7
This advertising is funny or clever	93.3
One of the best ads I've seen recently	93.3

Table 2
Factor Analysis (All Off-Air Responses)

Factor/item	Loading
<i>Factor I (38.0%)</i>	
<i>Informational/Consideration Frame</i>	
Helps me to find the product that I want	.935
I learned a lot from this advertising	.845
A brand I would consider buying	.838
Overall brand rating	.834
Played back advertising message	.820
Played back main point of advertising message	.820
In poor taste	-.896
Doesn't give any facts, just creates an image	-.894
Insults the intelligence of the average consumer	-.886
<i>Factor II (22.3%)</i>	
<i>Positive Commercial Reaction</i>	
One of the best ads I've seen recently	.953
I find this advertising artistic	.935
This advertising is enjoyable	.926
Liking scale	.862
This advertising is boring	-.892
<i>Factor III (16.7%)</i>	
<i>Persuasion</i>	
Purchase intent scale (DWB)	.832
This advertising is funny or clever	.710
Brand most likely to buy	.397
Claimed recall from full set of cues	-.760
<i>Factor IV (6.1%)</i>	
<i>Brand Recall</i>	
Recall brand from Product Category Cue	.730

markets were not typical of national sales and that, had national sales data been available, the on-air test would have done a better job of predicting them.

Another complicating factor may be the effects of program environment. Early work by Sonia Yuspeh, Gert Assmus, and more recently by David Lloyd, Kevin Clancy, and others documents the fact that the program in which a commercial appears can have a significant effect on its performance. Program environment was a constant in the off-air tests. The same program was used for all off-air tests. However, program environment varied for the on-air tests depending upon what was available.

The conclusion on this issue, therefore, is that the off-air methods did a reasonably good job predicting sales performance in the geographic areas in which the testing was done. No firm conclusions are possible about the validity of on-air testing.

Pre/post versus Post Only. A second methods issue is whether the use of a pre/post design is preferable to use of a post-only (matched group) design. One of the advantages of pre/post designs is that each person acts as his or her own control, thereby making the job of detecting changes stemming from exposure easier. Significance bands are narrower, and there are financial advantages as well.

On the other hand, it is easier to simulate normal exposure with the post-only design and to avoid sensitizing respondents to the fact that they are in an experiment designed to measure how their attitudes are changed by exposure to a commercial. But the price paid for this advantage is the need for accurate matching of the exposed and unexposed groups. If this is not accomplished, any observed differences

Table 3
Factor Analysis (Pre/Post Cells Only)

Item	Loading
<i>Factor I (40.4%)</i>	
<i>Informational/Consideration Frame</i>	
Helps me to find the product that I want	.923
I learned a lot from this advertising	.910
A brand I would consider buying	.900
Played back advertising message	.888
Told me something new about the product	.884
Change in consideration frame	.862
Change in top-of-mind awareness	.850
Doesn't give any facts, just creates an image	-.935
In poor taste	-.934
Insults the intelligence of the average consumer	-.874
<i>Factor II (20.7%)</i>	
<i>Persuasion/Liking</i>	
Overall Brand Rating	.919
Liking scale	.871
One of the best ads I've seen recently	.812
I find this advertising artistic	.805
Change in constant sum	.774
This advertising is enjoyable	.741
Purchase intent scale (DWB)	.737

may relate to differences in audience composition rather than to the effects of commercial exposure.

A number of years ago, Neil DuBois did a series of careful experiments at Foote, Cone & Belding during which pre/post and matched-group designs were half sampled for a large sample of commercials. He established the fact that the choice between these two designs can and often does have a significant effect on results. However, he had no outside criterion to settle the issue of which is the *better* design.

Figure 12 summarizes the results obtained during the ARF research. They are based on comparisons of the off-air pre/post cells with the off-air post-only cells.

On the surface the post-only design has a distinct edge in most areas. However, it is hard to generalize from these results to all copy tests because, as it turned out in retrospect, the audiences for the pairs of commercials researched in the ARF tests were very well matched in almost every case. This, of course, favors the post-only design. But whether this can be expected to be a regular occurrence depends heavily upon both respondent selection and matching procedures and, of course, on sample sizes.

An extensive analysis was done to see what the effects of post matching might be. Statistical adjustments were made to bring the audiences for each pair of commercials into the closest

Figure 11

On-Air vs. Off-Air (50% Confidence Level)

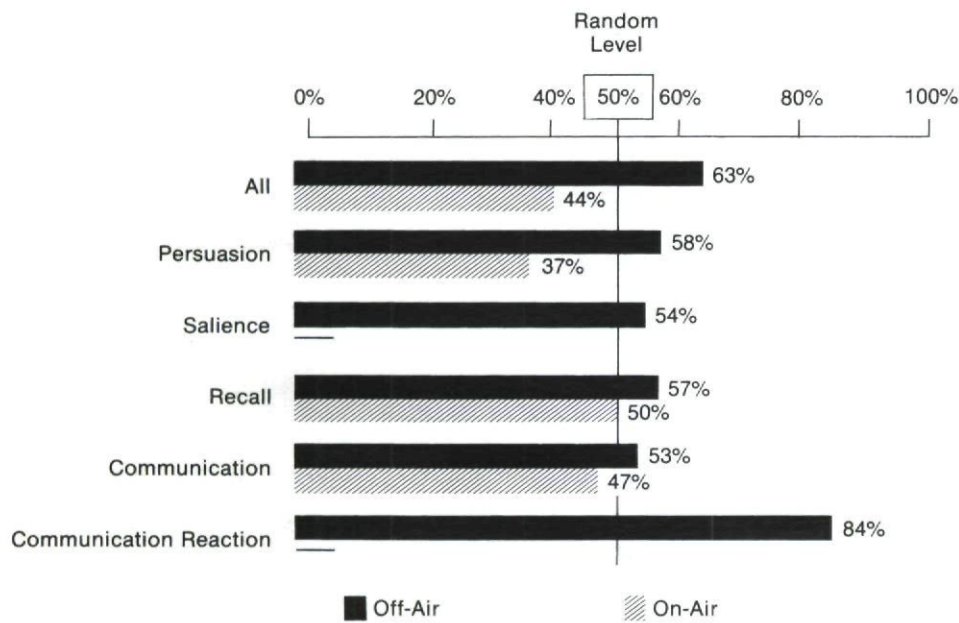
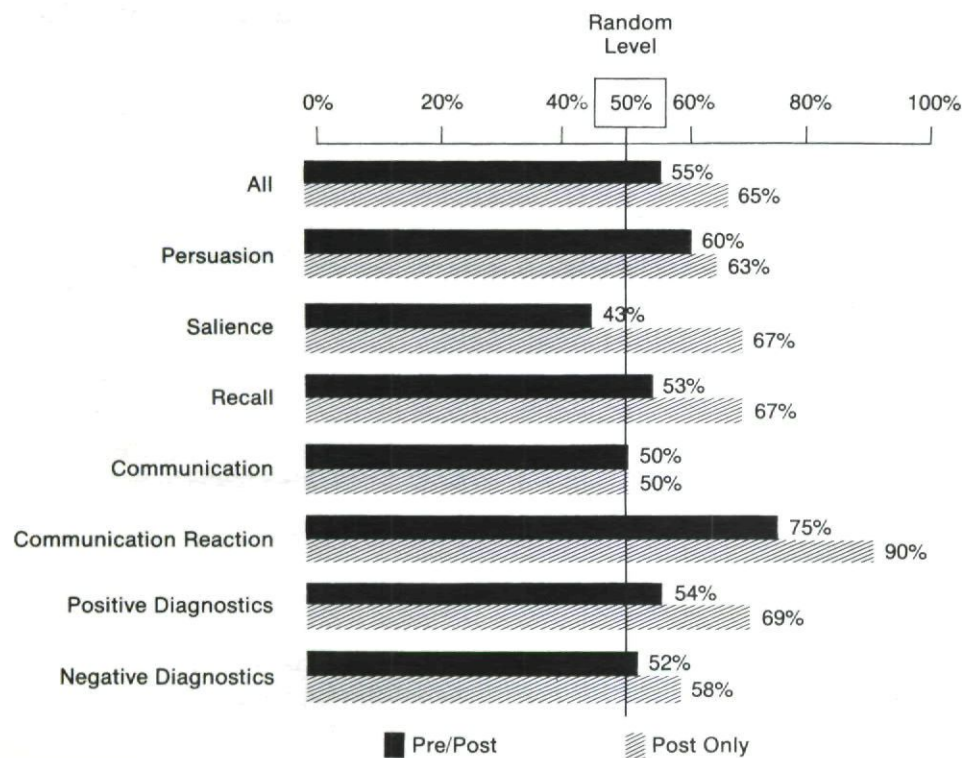


Figure 12

Pre/Post vs. Post Only* (50% Confidence Level)



* Based on off-air cells only

possible alignment. However, while the adjustments consistently moved scores in the right directions, the average magnitude of the adjustments across measures was only 0.3 percentage points and thus made little difference in the results. So no further work was done with adjusted data.

Summing up, the findings here suggest an edge for the post-only design. However, the maintenance of that edge depends upon the proficiency of the matching procedures used.

Single Exposure versus Multiple Exposure (Re-exposure). There are advocates for each of these design alternatives as well. The multiple-exposure design raises the intensity of the advertising copy as a stimulus and thus gives the researcher a better chance of obtaining a measurable impact. By the same token it allows the gathering of more detailed diagnostic information. However, people who favor the single-exposure design point out that it is more realistic in that it is closer to the way in which people are normally exposed to commercials. And, therefore, they believe its measurements are more likely to be related to subsequent sales performance.

Figure 13 summarizes results for the two off-air cells that differ only with respect to single-exposure/multiple-exposure issues. By way of reminder, the re-exposure design involves exposing the commercial in a program environment, asking people their opinions of the program, and then exposing people to the commercial again—this time without the program. The questioning sequence that follows is then exactly the same as in the single-exposure design.

On an overall basis there is not much to choose from between these two sets of results. The

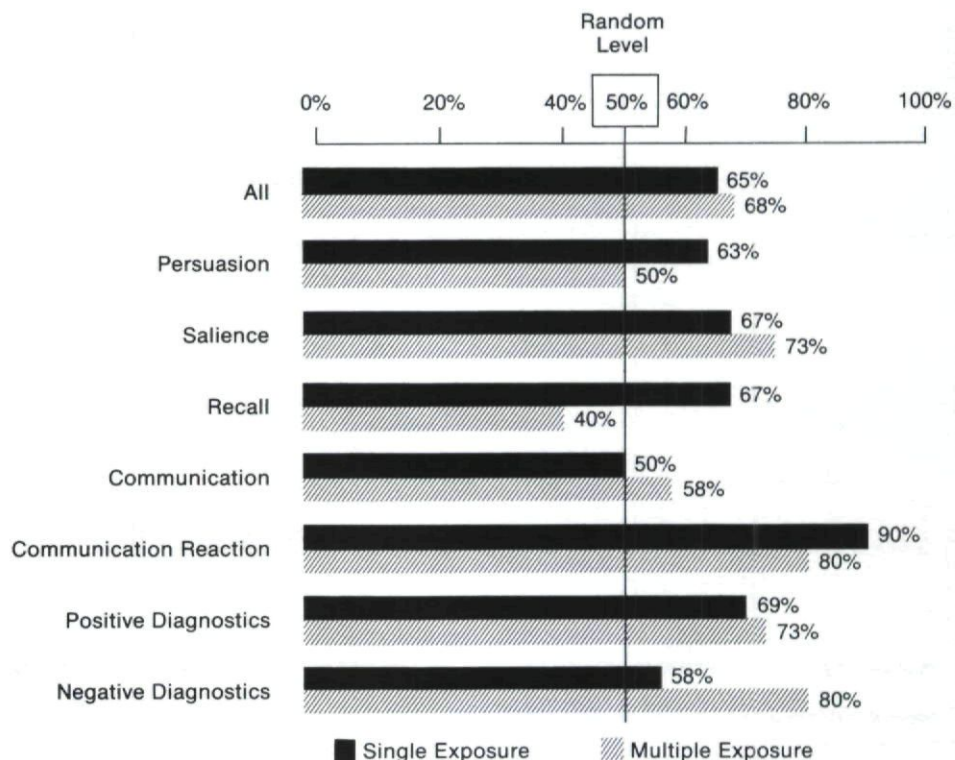
multiple-exposure design appears to improve the predictive powers of the diagnostic measures. However, recall, persuasion, and liking all showed greater predictive accuracy in the single-exposure design. So based on these data, the choice between designs depends largely on which specific measures form the basis for action standards. One hypothesis worth testing in the future is that the best choice of all may be a compromise design in which recall, persuasion, and liking measures are made after one exposure and diagnostic measures after re-exposure.

Scale Scoring. A supplementary question, examined in the course of tabulation, was whether some methods of scale scoring were superior to others. There are many ways in which scales might be scored—mean ratings of scales scored in a linear fashion, top-box ratings, top-two-box ratings, bottom-box ratings, and logarithmic or exponential weights—to name a few that have been used. Figure 14 summarizes results for the two most popular methods of scale scoring—arithmetic means and top-box ratings.

Overall the differences in scoring systems do not produce significantly different predictive accuracy in this study. Consequently, it is probably best to fall back on the rule of thumb that when samples are small the use of arithmetic means is preferred because they offer greater reliability. But when samples are large the use of top-box scoring is preferred because it offers greater sensitivity and is more strongly related to behavior. In the ARF study there are two samples—a large one of 15,000 respondents and a small one of 10 commercials. The fractional predictive superiority of mean ratings (69.3 percent to 66.7 percent) suggests that in this instance the number

Figure 13

Single vs. Multiple Exposure (Re-exposure)* (50% Confidence Level)



* Based on off-air cells only

of commercials may be the dominant factor.

Discussion of Findings and Implications for Users of Copy Testing

One standard for evaluating the utility of any research study is to judge how well it fits in with prevailing expert opinion—in this case expert opinion on the aspects of copy that affect sales and expert opinion on how copy is best tested when clear-cut sales results cannot be obtained. If the results of the research do not fit well, then either the study is flawed or experts need to modify their opinions in the light of fresh evidence.

The trouble with applying such a standard to research on copy testing is that there are a *variety*

of expert opinions as evidenced by the variety of commercial copy-testing systems offered. Moreover, the type of study summarized here has never been conducted before. So the opinions of copy-testing experts are based on a variety of research approaches to the question of how copy-test results relate to actual sales effects. It may well be true, for example, that other validation efforts relate more to short-term copy effects whereas the ARF Copy Research Validity Project is concerned with year-long effects.

Copy Testing Works. One thing that most copy-research professionals agree upon is that copy testing works—that it does relate to sales. And the ARF study confirms that fact. Copy testing is definitely helpful in

identifying commercials known to be generating incremental sales. Moreover, copy differences can be important—as evidenced by the fact that we found pairs of commercials that under rigorous testing, were producing large sales differences.

Multiple Measures Are Needed. The study also confirms the PACT Principles endorsed by the research directors of the 21 largest advertising agencies in the United States earlier in the 1980s. The principle that states that advertising works on a number of levels and, therefore, that no single measure is adequate to measure the effectiveness of copy seems to receive strong confirmation here. Multiple measurements are necessary for a full evaluation of copy effectiveness.

All Types of Copy Testing Measures in Common Practice Have Predictive Value. It was also reassuring to find that, within each of the types of measures in common use today, there is at least one specific measure showing a positive relationship to sales performance. Persuasion, recall, copy playback, brand salience, and commercial reaction all can play a role in the effectiveness of the copy.

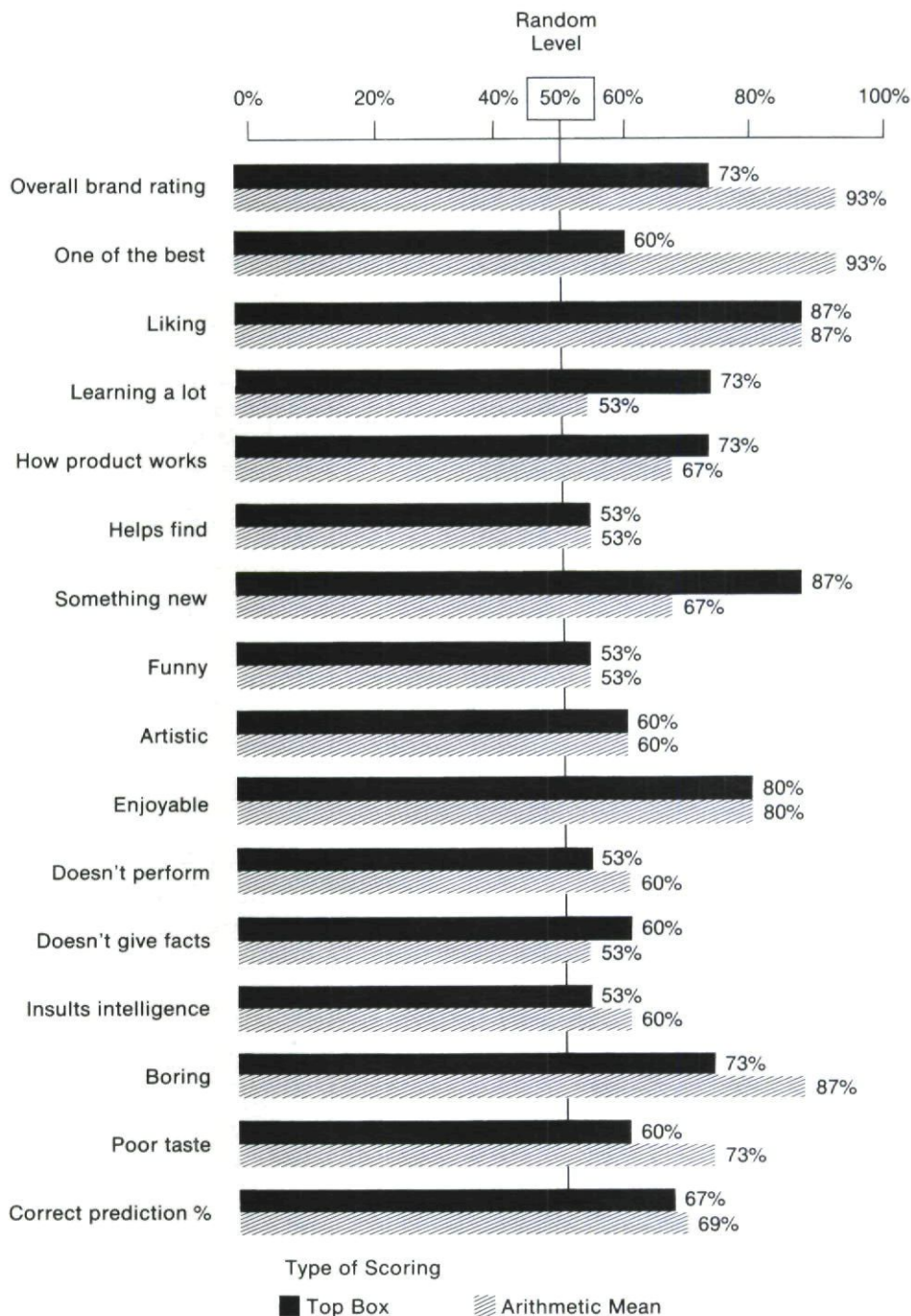
Some Surprises within Types

Within types of measures, the specific measures that were the best predictors of sales were generally in line with conventional wisdom (e.g., “main point” in the communications category and “top of mind” in the salience category). However, there were also some mild surprises. For example, in the recall area results suggested minimal probing. More pointedly, correct recall of the brand advertised (following only product category cue) predicted sales best of the recall-type measures. In the persuasion

area, Overall Brand Rating had a somewhat better record than the pre/post Constant Sum Scale or the well-known Definitely Will Buy Scale.

The latter point deserves further discussion because of some confusion surrounding the definition of persuasion. In some instances definitions of persuasion

Figure 14
Effects of Scale Scoring* (% Correct Predictions)



* Based on off-air cells only

have been limited to measures of choice. However, for purposes of the ARF research, persuasion was defined as "increasing the probability of purchase by making people more favorably predisposed to the advertising brand." This can be measured in a number of ways. The Arrowhead #9 study mentioned earlier investigated 13 persuasion measures and found them all to be significantly related to each other. Likewise, as noted earlier in this article, the constant sum and purchase intent measures factor together. However, there are subtle differences between them. When the competitive choice set is well defined, the constant sum measure is a very attractive measure. However, in situations in which an industry leader is interested in expansion of its category or where a significant portion of brand sales are likely to come from products beyond the most immediate competitors, the Overall Brand Rating may be a better measure. Of the five brands represented in the ARF study at least one brand, and arguably two, do not have a well-defined choice set.

Post-Only Did Especially Well in This Project. In a related issue the Post-Only design fared better than some had expected in relation to the Pre/Post design. However, caution is urged in generalizing from this finding. It is highly likely that the success of the Pre/Post design depends heavily upon how well its purposes are disguised from the respondent—the minimization of the transparency bias. Also it has some economic advantages not dealt with in this research.

Off-Air Also Did Well. An encouraging finding from an economic standpoint was the strong showing of off-air laboratory approaches. The buying of expensive media time does not appear to be a prerequisite to sound

copy testing. Nor does money necessarily have to be spent on costly cut-ins and cut-outs. It should be possible to do acceptable developmental work not only in commercial laboratories but in university laboratory settings.

No Methods or Measures Can Be Rejected from This Study. It should be emphasized that no copy-testing methods nor measures can be rejected as a result of the ARF study. It was not designed to do so. The hypothesis of no difference, the "null hypothesis," can never be proved, certainly not on the basis of 5 pairs of commercials. It is always possible that, with other commercials, other brands, other markets, other time periods, and different question phrasing, relationships would be found between sales and the measures which did less well in this research. On the other hand, measures that *do* show strong relationships to sales performance in this study are certainly worth the attention of the copy-research community.

Project Strongly Suggests More Attention to Likability

Undoubtedly the most surprising finding in the study was the strong relationship found to exist between the likability of the copy and its effects on sales. Successive generations of applied researchers, dating back to Alfred Politz, had learned that they should avoid questions that would turn the respondent into an expert. Also it was generally believed that the function of advertising was to sell and that whether the copy was likable or not had little to do with it. And there were the well-known examples of ads that most people considered to be abrasive which were nevertheless reported to be

effective generators of sales. But, while there may be exceptions to any rule, the ARF study strongly suggests that ads that are liked outsell those that are not.

Other support for the value of likability appeared at the July 1990 ARF Copy Research Workshop. Jim Spaeth of Viewfacts, Inc. reported that liking was an excellent predictor of sales generated from five pairs of *TIME* magazine commercials concerning a variety of subscription offers. Gordon Brown of Millward Brown, Inc. cited a high correlation between likability and his awareness index which, in turn, he finds to be the strongest correlate of sales increases reflected in tracking studies. He has three years of data and a large number of cases.

Alex Biel of the Center for Research and Development summarized an Ogilvy study spanning 57 products in 11 categories that showed that likable ads were twice as persuasive as the average ad. So the case for likability is documented from other applied sources.

Academic researchers have also shown evidence, in the context of research on "Attitude toward the Ad" and its effects on brand attitudes and purchase intentions, that likability is an important mediator of message effects. Esther Thorson, in a 1990 review of "Consumer Processing of Advertising" for *Current Issues & Research in Advertising*, sums them up by saying "These studies leave little doubt that the individual's response to the ad itself is a powerful predictor of ad impact."

But exactly what is likability? Is it the same as humor? By no means. The evidence from the Validity Project suggests that likable ads are just as apt to be informative as they are to be entertaining ads. And even ads *designed* to be entertaining may not

be considered likable by consumers.

Fortunately, an in-depth study of likability by Alex Biel was also presented at the 1990 ARF Copy Research Workshop. He found that it had five dimensions which he labeled "Ingenuity," "Meaningfulness," "Energy," "Warmth," and "Rubs the Wrong Way." So likability is a complex concept. That study, together with the promising academic research referenced at the end of this article, point the way to additional productive research on the ways in which advertising can influence sales.

What are current theories on how likability exerts its influence on sales? *USA Today*, in its August 29, 1990, issue, quoted Arie Kopelman, president of Chanel, Inc., as saying, "If a commercial is interesting you could watch it a thousand times. If it's obvious and boring after you see it twice you could put a shoe through the television set."

Biel offers the following hypotheses:

- (1) Commercials that are liked get more exposure (also the Kopelman hypothesis).
- (2) Commercials are brand personality attributes and affect sales through their overall contribution to the reputation of the products.
- (3) Ads that are liked are given more mental processing (liking is a mediator).
- (4) Liking is a "gatekeeper" to whether or not the ad is processed at all. (Liking is a moderator.)
- (5) There is less counterarguing against ads that are liked.
- (6) Liking engenders trust (source credibility).
- (7) Liking the commercial translates directly to liking the brand (emotional ruboff).
- (8) Liking evokes a gratitude response. Consumers buy the

product to reward the advertiser for likable advertising.

As Mike Naples, president of the ARF, has observed, one function of any study like this one is to point the way toward further potentially productive research. Hopefully, it is clear that that objective has been abundantly achieved. In any event, given the amounts of time, energy, and money required for a study of this type, plus the difficulties of locating eligible pairs (no one has come forward with one despite open invitations in April and July 1990 to do so), the study is unlikely to be replicated in the foreseeable future. However, if indirectly it results in commercials that are judged by consumers to be more likable and if that, in turn, softens criticism of the advertising industry and stimulates further research on likability, the tremendous amounts of time, energy, and money expended by the many people connected with the ARF Copy Research Validity Project will all have been worthwhile.

As a concluding note, the study in no way implies that likability should be considered as a stand-alone measure of copy effectiveness. Persuasion and recall justifiably remain as important copy-testing measures and are likely to remain primary evaluative measures in the foreseeable future. ■

References

- Aaker, D. A.; D. M. Stayman; and M. R. Hagerty. "Warmth in Advertising: Measurement, Impact and Sequence." *Journal of Consumer Research* 12, 4 (1986): 365-81.
- Alsop, R. "TV Ads that are Likeable get Plus Rating for Persuasiveness." *Wall Street Journal*, February 29, 1986.
- Alwitt, L. "Components of Ad Likeability." Paper presented at the Second Walter N. Stellner Symposium on Uses of Cognitive Psychology in Advertising and Marketing. Champaign-Urbana, IL, 1987.
- Assmus, G. "The Empirical Investigation Into the Perception of Vehicle Source Effect." *Journal of Advertising* 7, 1 (1978): 4-10.
- Batra, R. "Affective Advertising: Role, Processes and Measurement." In *The Role of Affect in Consumer Behavior: Emerging Theories and Applications*. Lexington, MA: D.C. Heath, 1984.
- , and M. B. Holbrook. "Development of a Set of Scales to Measure Affective Responses to Advertising." *Journal of Consumer Research* 14, 3 (1987): 404-20.
- , and M. L. Ray. "Affective Responses Mediating Acceptance of Advertising." *Journal of Consumer Research* 13, 2 (1986): 234-49.
- , and D. Stephens. "Attitudinal Effects of Ad-Evoked Affective Responses." Working Paper, Columbia University, NY, 1987.
- Burke, M., and J. Edell. "The Impact of Feelings on Ad-Based Affect and Cognition." *Journal of Marketing Research* 26, 1 (1989): 69-83.
- Day, R. L. "Revisiting the Rough/Finished Issue in Advertisement Pretesting: A Practitioner's Viewpoint." *Marketing Research: A Magazine of Management Applications*, September 1990.
- Edell, J. A., and M. C. Burke. "The Moderating Effect of Attitude Toward an Ad on Ad Effect-

- tiveness Under Different Processing Conditions." *Advances in Consumer Research* 11 (1984): 644-49.
- Gardner, M. P. "Does Attitude Toward the Ad Effect Brand Attitude Under a Brand Evaluation Set? *Journal of Marketing Research* 22, 2 (1985): 192-98.
- Gelb, B. D., and C. M. Pickett. "Attitude-Toward-the-Ad: Links to Humor and to Advertising Effectiveness." *Journal of Advertising* 12, 2 (1983): 34-42.
- Gresham, L. G., and T. A. Shimp. "Attitude Toward the Advertisement and Brand Attitudes: A Classical Conditioning Perspective." *Journal of Advertising* 14, 1 (1985): 10-17.
- Haley, R. I., and P. B. Case. "Testing Thirteen Attitude Scales for Agreement and Brand Discrimination." *Journal of Marketing* 43, 4 (1979): 20-32.
- ; J. Richardson; and B. M. Baldwin. "The Effects of Non-verbal Communication on Television Advertising." *Journal of Advertising Research* 24, 4 (1984): 11-18.
- Homer, P. M. "The Mediating Role of Attitude Toward the Ad: Some Additional Evidence." *Journal of Marketing Research* 27, 1 (1990): 78-86.
- Lloyd, D. W., and K. J. Clancy. "Television Program Involvement and Advertising Response: Some Unsettling Implications for Copy Research." Paper Submitted to the *Journal of Consumer Marketing*, 1990.
- . "CPM's Versus CPMI's: Implications for Media Planning Based on New Evidence Regarding TV 'Program Environment' Effects." Paper dated February 1990.
- Lutz, R. J. "Affective and Cognitive Antecedents of Attitude Toward the Ad: A Conceptual Framework." In *Psychological Processes and Advertising Effects*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1985.
- ; Mackenzie, S. B.; and G. E. Belch. "Attitude Toward the Ad at a Mediator of Advertising Effectiveness." *Advances in Consumer Research* 10 (1983): 532-39.
- MacInnis, D. J., and B. J. Jaworski. "Information Processing from Advertisements: Toward and Integrative Framework." *Journal of Marketing* 53, 4 (1989): 1-23.
- Mackenzie, S. B., and R. J. Lutz. "The Empirical Examination of the Structural Antecedents of Attitudes Toward the Ad in an Advertising Pretesting Context." *Journal of Marketing* 53, 2 (1989): 48-65.
- ; ———; and G. E. Belch. "The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations." *Journal of Marketing Research* 23, 2 (1986): 130-43.
- Madden, T. J.; C. T. Allen; and J. L. Twible. "Attitude Toward the Ad: An Assessment of Diverse Measurement Indices Under Different Processing Sets." *Journal of Marketing Research* 25, 3 (1988): 242-52.
- ; K. Debevec; and J. L. Twible. "Assessing the Effects of Attitude-Toward-the-Ad on Brand Attitudes: A Multi-Trait-Multimethod Design." Working Paper 84-15. Amherst, MA: University of Massachusetts, 1984.
- ; W. R. Dillion; and J. L. Twible. "Construct Validity of Attitude-Toward-the-Ad: An Assessment of Convergent/Discriminant Dimensions." Working Paper 84-15. Amherst, MA: University of Massachusetts, 1984.
- Messmer, D. J. "Repetition and Attitudinal Discrepancy Effects on the Affective Response to Television Advertising." *Journal of Business Research* 7, (1979): 75-93.
- Mitchell, A. A., and J. Olson. "Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitudes?" *Journal of Marketing Research* 19, 18 (1981): 318-32.
- . "The Effect of Verbal and Visual Components of Advertisements on Brand Attitudes and Attitude Toward the Advertisement." *Journal of Consumer Research* 13, 1 (1986): 12-24.
- Moore, D. L., and J. W. Hutchinson. "The Influence of Affective Reactions to Advertisements: Direct and Indirect Mechanisms of Attitude Change." In *Psychological Processes and Advertising Effects*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1985.
- . "The Effects of Ad Affect on Advertising Effectiveness." *Advances in Consumer Research* 10 (1983): 526-31.
- Muehling, D. D. "Comparative Advertising: The Influence of Attitude-Toward-The-Ad on Brand Evaluation." *Journal of Advertising* 16, 4 (1987): 43-49.
- ; J. J. Stoltman; and S. Mishra. "An Examination of the Cognitive Antecedents of Attitude-Toward-The-Ad." *Current Issues and Research in Advertising* 12, 1 & 2 (1990): 95-117.

Percy, L. "Perspectives on Measuring Attitudes Toward the Ad." In *Marketing Communications—Theory and Research*. Chicago: American Marketing Association, 1985.

Petty, R. E., and J. T. Cacioppo. "Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Response." *Journal of Personality and Social Psychology* 37 (1979): 1915–26.

———, and ———. "The Elaboration Likelihood Model of Per-

suasion." In *Advances in Experimental Society Psychology* 19. New York: Academic Press, 1986.

———; ———; and D. Schumann. "Central and Peripheral Routes to Advertising Effectiveness." *Journal of Consumer Research* 10, 2 (1983): 135–146.

Shimp, T. A. "Attitude Towards the Ad as a Mediator of Consumer Brand Choice." *Journal of Advertising* 10, 2 (1981): 9–15.

Stayman, D. M., and D. A. Aaker. "Are All the Effects of

Ad-Induced Feelings Mediated by A_{AD} ?" *Journal of Consumer Research* 15, 4 (1988): 368–73.

Thornson, E. "Consumer Processing of Advertising." *Current Issues and Research in Advertising* 12, 1 & 2 (1990): 197–230.

Yuspeh, S. "The Medium Versus the Message: Effects of Program Environment on the Performance of Commercials." Presentation to the Tenth Attitudes Research Conference, American Marketing Association, 1979.

Appendix

Correlations between Types of Measures

Measures	Persuasion	Recall	Communication	Commercial reaction	Positive diagnostics	Negative diagnostics	Saliency
Persuasion	—	.733	.200	.867	.733	-.690	.552
Recall	.733	—	-.067	.600	.733	-.552	.276
Communication	.200	-.067	—	.333	.200	-.138	.000
Commercial reaction	.867	.600	.333	—	.867	-.552	.414
Positive diagnostics	.733	.733	.200	.867	—	-.552	.276
Negative diagnostics	-.690	-.552	-.138	-.552	-.552	—	-.214
Saliency	.552	.276	.000	.414	.276	-.214	—

Correlations between Methods

	Pre/Post	Post Only	Re-Exposure
Pre/Post	—	.684	.532
Post Only	.684	—	.564
Re-Exposure	.532	.564	—



Copyright of Journal of Advertising Research is the property of World Advertising Research Center Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.