

HOW COPY TESTING WORKS

Mapes and Ross Copy Test Validation Experience

by
Harold L. Ross
Chairman
Mapes and Ross
for

The ARF Brand Equity Research Day

How Copy Testing Works

Mapes and Ross Copy Test Validation Experience

In the short time available today, I would like to accomplish two things. The first is to share with you four examples of formal sales validation experience gathered by Mapes and Ross over the last ten years. In covering this experience, the emphasis will be on the learning that was gained that should be borne in mind in investigations of the association between attitudinal data and actual data.

The second part of my talk addresses the IRI study of validity of copy testing measures and its conclusions. As you will see, prior learning on this crucial subject was essentially ignored by the IRI study. Furthermore, a closer look at this study reveals that the data cited do not support their over-publicized conclusion regarding the validity of copy testing measures.

- REVIEW 10 YEARS OF SALES VALIDATION EXPERIENCE AND WHAT WE'VE LEARNED FROM IT.
- REVIEW THE ANALYTICAL PROCEDURE AND CONCLUSIONS OF THE IRI STUDY RE COPY TESTING.

At the outset, it's important to make the point that informal validation of copy test results is a frequent occurrence for many advertisers. The difference between formal and informal validation, from our perspective, is the extent to which these investigations come to grips with the many complexities involved in singling out the effect of advertising from the myriad of other marketing influences.

Informal validation tends to be simplistic; if the advertising is good and there are no obvious major changes in the marketing environment, sales should be good.

Formal validation is much more scientific, but is also time consuming and extremely costly.

In terms of specific test designs, there are essential two different approaches that can be employed.

- DETERMINE BUYING RATES FOR COMMERCIALS WITH HIGH PERFORMANCE SCORES vs THOSE WITH LOWER PERFORMANCE SCORES.
- DETERMINE BUYING RATES AMONG THOSE EXPOSED TO COMMERCIALS WHO ARE PERSUADED AND/OR RECALL THE COMMERCIAL vs THOSE EXPOSED WHO WERE NOT PERSUADED AND/OR DO NOT RECALL THE COMMERCIAL.

Our formal validation studies have employed both test designs. Both have documented clearly the same positive relationship between certain copy testing measures and actual sales in the marketplace.

Our first formal study of the association between our copy test measure and actual sales still stands as the most substantial copy test validation study yet published. This study involved follow-up interviews with two groups of respondents who had participated in standard Mapes and Ross day-after-exposure copy tests. One group comprised the test group (exposed to advertising for test brands). The other group, the control group, was identical, except they were exposed to advertising for other brands - but not the test brands. Buying rates of those exposed to advertising were compared with those not exposed for 142 commercials for established brands. This study was quite extensive, based on 2,241 exposed respondents for a total of 7,283 category purchase opportunities

1982 - MAPES AND ROSS VALIDATION STUDY -
DAY-AFTER RECALL vs DAY-AFTER PERSUASION- AN ANSWER

- BROAD BASED LONGTITUDINAL STUDY
- EXPOSED - UNEXPOSED SAMPLES
- ESTABLISHED BRANDS ONLY

The data yielded a wealth of information regarding the relationship between our copy test performance measures and actual sales. In the context of today's talk, let me emphasize four major conclusions pointed up by these data.

WHAT WAS LEARNED/CONFIRMED:

- EXPOSED VIEWERS BOUGHT MORE THAN UNEXPOSED VIEWERS.
- INCREASE IN BUYING PRIMARILY DUE TO THOSE PERSUADED BY ADVERTISING (WHOSE BUYING RATE WAS 3 TIMES AS HIGH AS NON-PERSUADED VIEWERS.
- DAY-AFTER PROVEN RECALL, BY ITSELF, IS A POOR MEASURE OF A COMMERCIALS EFFECT ON CONSUMER PURCHASE.
- THE SLIGHT MEASURABLE EFFECT ON PURCHASE OF DAY-AFTER PROVEN RECALL PERFORMANCE IS AMONG THOSE WHO ALREADY PREFERRED THE BRAND PRIOR TO EXPOSURE.

(A detailed description of this study appeared in February/March 1982 Journal of Advertising Research, and was reprinted in the Journal of Advertising Research Classics: Eight Key Articles That Have Led Our Thinking.)

Another formal validation study that we are at liberty to discuss publicly employed the other approach. That is, it compared the in-market sales effectiveness of two different campaigns for National Dairy Board.

1989 NATIONAL DAIRY BOARD CASE STUDY
(PRESENTED AT ARF WORKSHOP 1989)

In this case, two different campaigns were measured for day-after-recall and day-after-persuasion. While no difference existed in terms of recall, there was a significant difference in terms of persuasion.

	<u>PERSUASION</u>	<u>RECALL</u>
CAMPAIGN 1	BELOW AVERAGE	AVERAGE
CAMPAIGN 2	ABOVE AVERAGE	AVERAGE

Both campaigns were aired and monitored by extensive tracking activity which measured both in-home consumption, and attitudinal impressions regarding consumption of this dairy product. Campaign 1, which achieved lower persuasion in the copy test, was aired first, for a period of several months. During this time, consumption was consistently slightly below the previous year.

When Campaign 1 was replaced by the more persuasive Campaign 2, in the same marketing environment and with no change in spending, this pattern was reversed and consumption increased dramatically. Additionally, positive change in attitudinal impressions indicated the source of this in-market change.

<u>DAIRY PRODUCT</u> <u>IN-HOME CONSUMPTION</u>	
<u>PERIOD</u>	<u>% CHANGE vs</u> <u>PREVIOUS YEAR</u>
1	- 1.6%
2	- 2.4
3	- 3.1
4	- 1.6
5	- 1.3
<hr/>	
6	+ 1.1
7	+ 1.9
8	+ 3.8
9	+ 6.2
10	+ 9.0

This extensive study corroborated the major findings of the earlier study and highlighted a crucial factor - changes in-market should be expected when there is a change in advertising effectiveness. In this case, above average persuasion showed very dramatic in-market results because it followed advertising that was below average!

NDB - WHAT WAS LEARNED/CONFIRMED:

- EXPOSURE TO THE MORE PERSUASIVE CAMPAIGN PRODUCED HIGHER BUYING BEHAVIOUR
- THERE WAS NO DIFFERENCE BETWEEN CAMPAIGNS IN TERMS OF DAY-AFTER RECALL.
- CHANGE IN BUYING BEHAVIOUR REFLECT CHANGE IN EFFECTIVENESS.

Our third formal example of in-market validation is proprietary, and, therefore, must be described in only general terms. it has been very broad in scope and has involved many commercials on varying levels of effectiveness. Incidentally, several of these in-market tests were conducted by IRI.

ON-GOING (9 YEARS) MULTIPLE COMMERCIALS FOR ESTABLISHED BRANDS - CONTROLLED MARKET TESTS.

- HIGH AND AVERAGE AND LOW PERSUASION.
- CATEGORIES CHARACTERISED BY HEAVY PROMOTIONAL SPENDING.

(PROPRIETARY)

The results of these studies have confirmed the patterns established by the other two studies.

WHAT WAS LEARNED/CONFIRMED - (PROPRIETARY)

- EXTREMELY HIGH CORRELATION BETWEEN DAY-AFTER PERSUASION SCORES AND ACTUAL BUYING BEHAVIOUR
- ADDITION OF DAY-AFTER RECALL SCORES IMPROVED CORRELATION VERY LITTLE.

Our most recent experience has involved an area frequently neglected by traditional copy test approaches - the influence of advertising on buying rates among current brand users. The validation approach used parallels that of our 1982 study - a longitudinal audit of purchases among exposed vs. unexposed respondents. In this case, however, the crucial distinction was the investigation of buying behaviour of both brand users and non-brand users in the test and control cells.

1990 - 1992 - SPECIAL STUDIES TO DETERMINE
EFFECT OF ADVERTISING AMONG CURRENT USERS.

- LONGITUDINAL AUDIT OF PURCHASES
BUYING RATES/UNITS.
- BRAND USERS COMPARED WITH NON-BRAND USERS

(PROPRIETARY)

Other studies have shown increased buying among current brand users may account for most of the sales produced by advertising. The sales results observed in our multiple studies have confirmed the importance of distinguishing between buying behaviour among brand users vs. non-brand users, despite the fact that most standard copy testing persuasion measures do not assess performance among current brand users. Commercials that perform poorly among non-brand users might do very well in increasing sales among brand users; however, special measures are required to identify when a commercial is effective among a brand user group.

These studies also indicated that day-after-recall can be an important measure among brand users.

WHAT WAS LEARNED/CONFIRMED:

- INCREASED SALES MAY COME FROM CURRENT BRAND
USERS AS WELL AS NON-USERS.
- SPECIAL PERSUASION MEASURES WERE SUCCESSFULLY
IDENTIFIED THAT PREDICT INCREASED
PURCHASE/ USAGE AMONG THOSE WHO ALREADY
PREFER/USE THE BRAND.
- IN SOME CASES, THE DAY-AFTER RECALL IS AN
IMPORTANT MEASURE AMONG BRAND
PREFERRERS/USERS.

To synthesize the learning developed from these many years (and dollars) of in-market validation, these are the four major cornerstones:

SUMMARY OF WHAT WAS LEARNED/CONFIRMED:

- DAY-AFTER PERSUASION IS PREDICTIVE OF SALES.
- FOR NON-BRAND USERS, DAY-AFTER PROVEN RECALL, BY ITSELF, IS NOT PREDICTIVE OF SALES.
- INCREASED SALES MAY COME FROM CURRENT BRAND USERS AS WELL AS NON-USERS (WHICH MAY NOT BE MEASURED BY TRADITIONAL PERSUASION MEASURES).
- CHANGES IN BUYING BEHAVIOUR DEPENDENT ON CHANGES IN EFFECTIVENESS OF CAMPAIGNS.

To amplify the final point regarding the importance of a change in advertising effectiveness, consider a situation in which two different brands each test two campaigns for persuasion with identical outcomes of +5 for Campaign I and + 10 for Campaign II. If brand A's advertising has historically been achieving persuasion levels of +5, then Campaign II should be expected to improve sales. Brand B, however, has had much more effective advertising in the past, with an average persuasion level of +15. In this case, both Campaigns I and II would be expected to have a negative outcome in terms of sales.

CHANGE IN ADVERTISING EFFECTIVENESS AND
CHANGE IN SALES

	BRAND A	BRAND B
PREVIOUS CAMPAIGN	+5	+15
TEST CAMPAIGN I	+5	+5
TEST CAMPAIGN II	+10	+10
<hr/>		
SALES OUTCOME	I HIGHER	I LOWER II LOWER

These findings are very significant as comparisons of copy test results and in-market behaviour are contemplated. This learning provides improved focus for analysis.

TO EVALUATE COPY TESTING RESULTS -

- LOOK AT BUYING BEHAVIOUR OF THOSE EXPOSED TO COMMERCIALS WITH DIFFERENT PERSUASION LEVELS.
- UNLESS SPECIAL MEASURES ARE USED TO ASSESS EFFECTIVENESS AMONG BRAND USERS, LOOK FOR DIFFERING PURCHASE BEHAVIOUR AMONG NON-BRAND USERS.
- BEAR IN MIND CHANGE IN BUYING BEHAVIOUR SHOULD DEPEND ON CHANGE IN CAMPAIGN EFFECTIVENESS.

Thus, meaningful findings would be forthcoming if appropriate measures, for appropriate segments of the market, for appropriate commercials (differing in effectiveness), are compared. Obviously the influence of weight, while an important advertising issue, has no place in this basic evaluation of copy testing performance measures.

CONTROLLED MARKET TEST
(BUYING RATES + UNITS)

	BRAND USERS	NON-BRAND USERS	TOTAL
TEST COMMERCIAL A (HIGH PERSUASION)			
TEST COMMERCIAL B (AVERAGE OR LOW PERSUASION)			
	NOT WEIGHT! RECALL AMONG BRAND USERS?		

Again, it must be borne in mind that standard copy testing measures do not address the persuasiveness of commercials among current brand users (although recall may be an important measure among this market segment).

<u>CONTROLLED MARKET TEST</u> (BUYING RATES + UNITS)			
	BRAND USERS	NON-BRAND USERS	TOTAL
TEST COMMERCIAL A (HIGH PERSUASION)	?		
TEST COMMERCIAL B (AVERAGE OR LOW PERSUASION)	?		
			NOT WEIGHT! RECALL AMONG BRAND USERS?

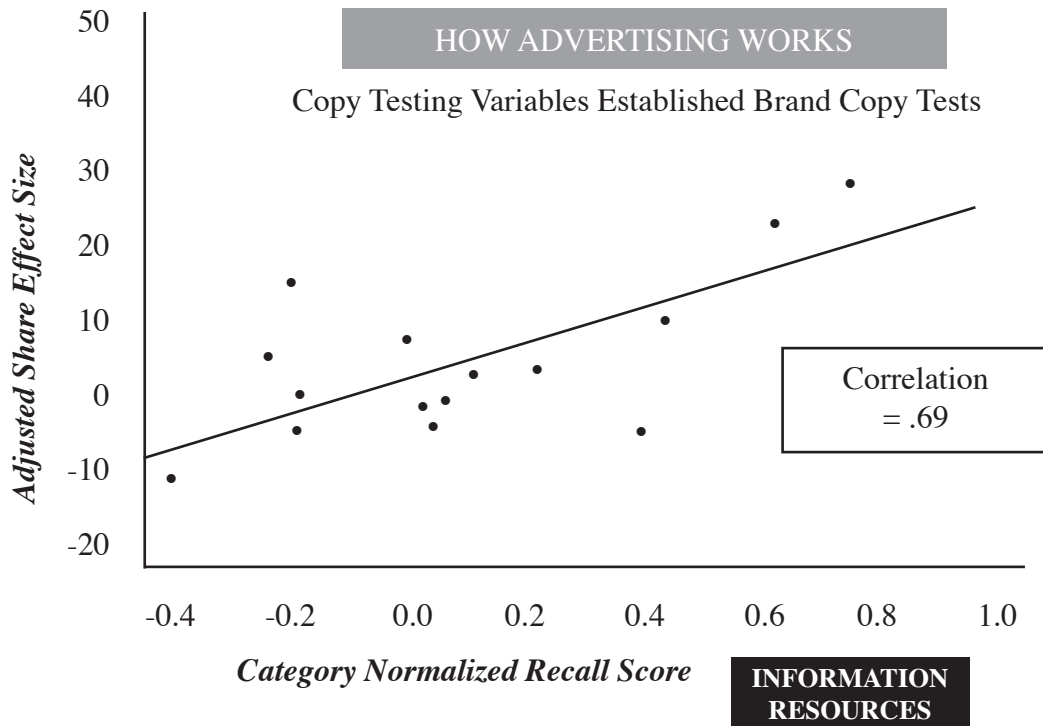
Before citing the actual data, it is important to be reminded that copy test measures from several different sources were combined in this analysis. Also, performance measures were indexed - sometimes on very low normative levels - which can be misleading at times.

(Measures that are statistically the same can end up with vastly different indices of performance). Also, definitions of the key criteria of “adjusted share of effect size” and “category normalized recall/persuasion scores” were not provided.

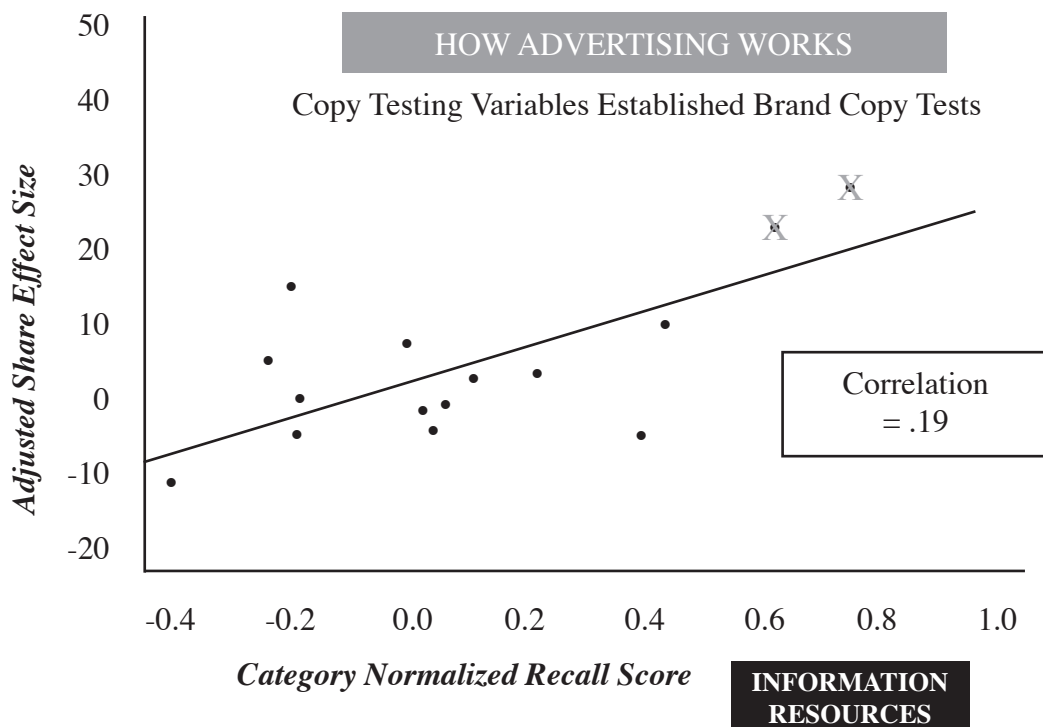
REVIEW OF IRI DATA

The first finding of the IRI presentation compared recall performance with “adjusted share effect size.” Despite an overall correlation of .69 for recall, the IRI conclusion was, “It is unlikely that there is a strong relationship between standard measures of TV commercial recall and persuasion for established brands and their ability to predict sales impact.”

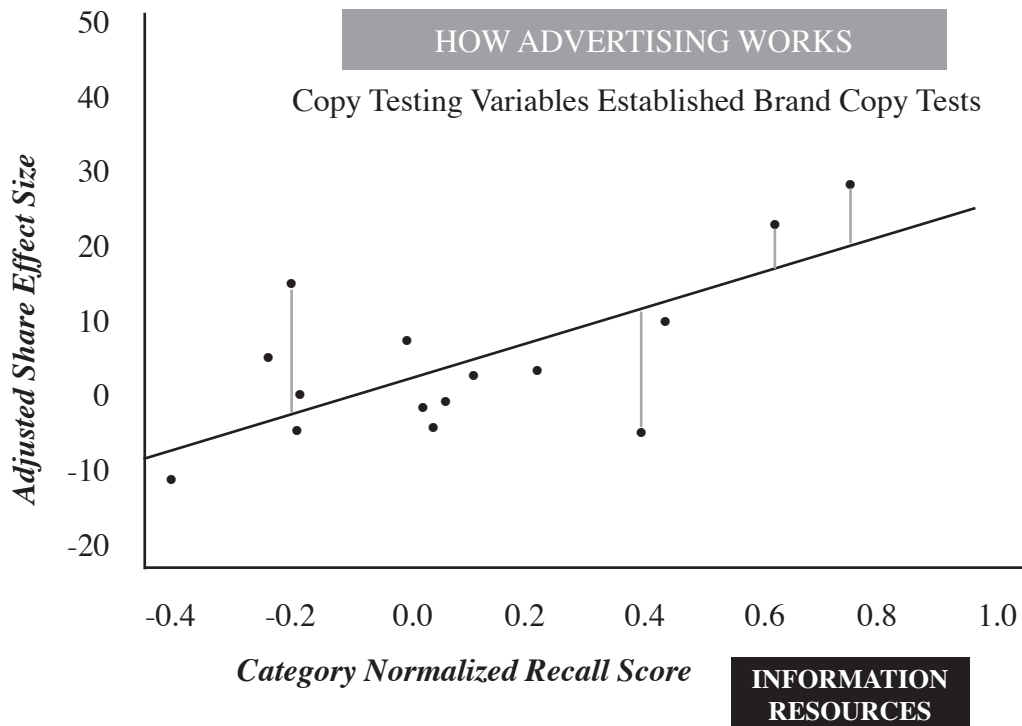
(Our data leads us to expect that this correlation is as high as it is because of the higher levels of persuasion among commercial recallers - as compared with non-recallers.)



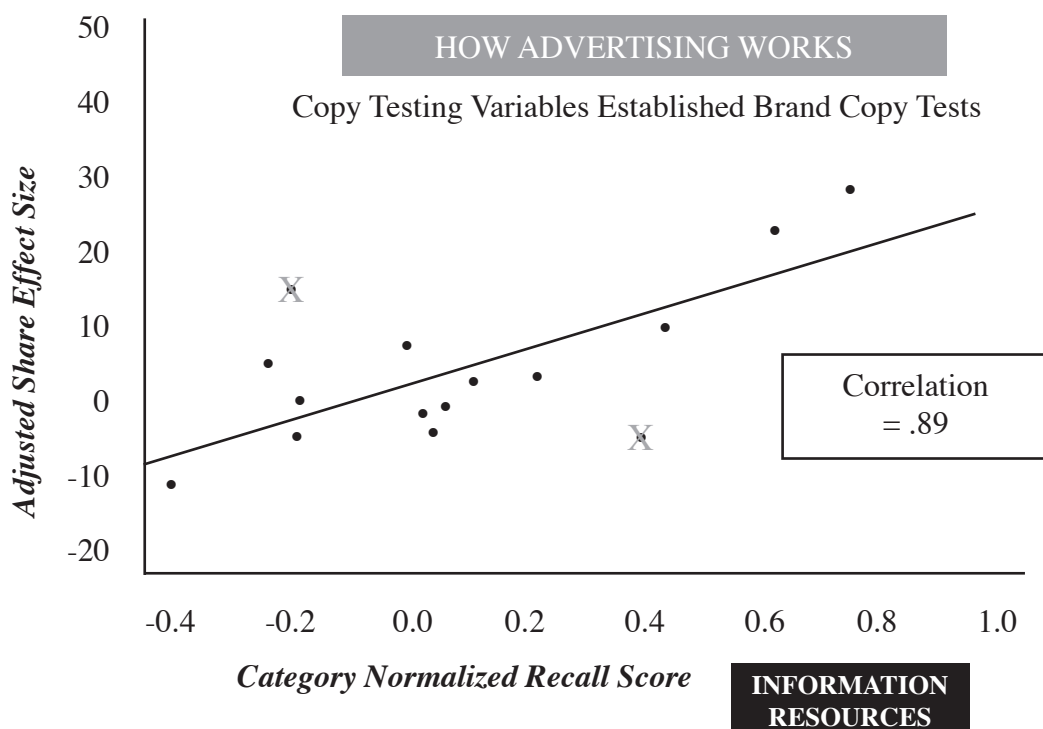
The explanation offered by IRI in justifying their conclusion is that, by eliminating the examples which showed the best relationship between advertising test results and share, the correlation was reduced to the point that it was no longer significant.



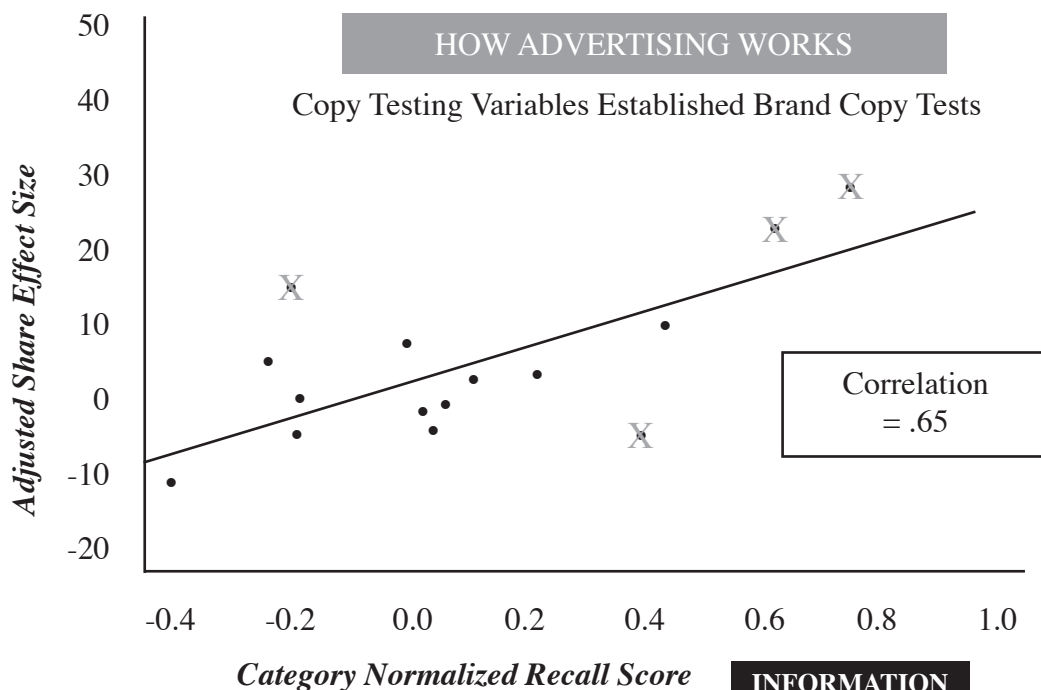
However, a look at the regression line based on all of the data suggested that any dismissal of “outliers” would most appropriately have eliminated other data points - rather than the ones they chose to ignore.



With these other “outliers” eliminated, the correlation becomes a breath stopping .89!

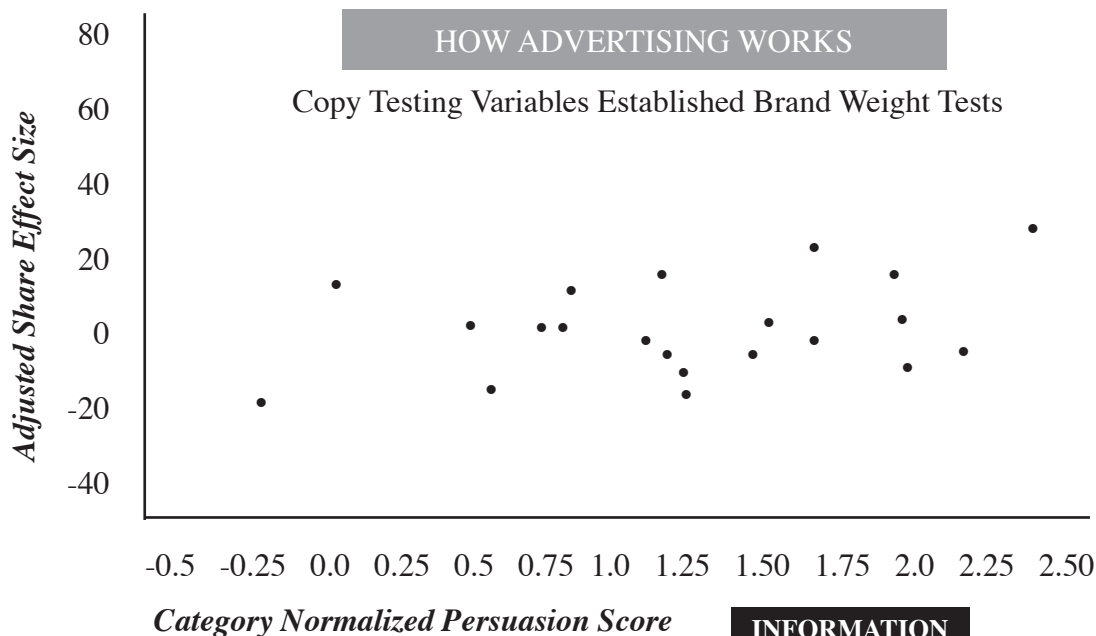


You may wonder what happens if all four of the nominees for removal were dropped from the data - it ends up very close to the very positive correlation indicated by including all data points!



**INFORMATION
RESOURCES**

Mapes and Ross' most serious disagreement with the IRI study concerns the fact that their conclusion that persuasion is tenuously related to sales impact is based solely on tests where the only variable was media weight. The report fails to clarify the bases for the analyses. Without a copy variable, only one persuasion score is available for comparison to two media weight/sales response levels rendering it impossible to draw any conclusion about the relationship of persuasion to sales change. If both a copy variable and media variable are present, they may attempt to draw a conclusion about the relationship of a single variable (persuasion) to sales when the other variables are affecting results is wholly inappropriate.



**INFORMATION
RESOURCES**

Having available in-market data of the kind provided by IRI, is a great boon to our industry.

It has shattered some of the conventional wisdom myths that afflict our industry.

But with this great source of information comes the burden of using it responsibly.

In the case of using the data to evaluate copy testing measurements, that responsibility can be shared by involving the people who know copy testing best - the suppliers who have spent their research careers developing stores of knowledge and providing improved services based on our collective learning.

The need for continued research must go on. Suppliers must continue to provide additional measures of effectiveness, such as our own work on brand users: and independent studies into how well systems based on these new measures perform must also be provided.

However, the interests of the advertising industry will only be served well when findings are subjected to more rigorous scrutiny and more exacting standards. The industry looks to the ARF to fulfill their responsibility, and Mapes and Ross pledges it's support to this effort.

Mapes & Ross

A division of Roy Morgan International

Roy Morgan International
176 Wall Street, Princeton, NJ, 08540-1583 USA
Phone (609) 924 8600 Fax: (609) 924 9208
www.roymorgan.com



Discover your edge

