PRINT ADVERTISING COPY TESTING: SOME PROBLEMS IN VALIDITY

Eduardus Tandelilin

ABSTRACTS

This article aims to develop an integral overview of the present status of the theory of the copy testing validity. It includes the presentation of the theoritical consideration or conceptual issues in copy testing, some empirical results or finding in copy testing, and finally, proposes an alternative conceptual framework of copy testing that a company can use to improve validity in copy testing especially in print copy testing.

Introduction

The effectiveness of any advertisement mainly depends on the prior preparations and external processes conducted before launching its final campaign. According to *Lilien, Kotler,* and *Moorthy* [1992), much of the effect of an advertising exposure depends on the creative quality of ad itself. But rating of the quality of the ad is extremely difficult. An advertisement may have very good aesthetic properties and win awards, but it may not do much for sales. On the other hand, an advertisement may seem crude and offensive, but it may be a major force behind sales *[Aaker* and *Day,* 1991].

Copy testing addresses issues related to the effectiveness of each advertisement. Although many studies show that copy testing has a problem surrounding measurement technique, there is little evidence that controversy about this technique has diminished [*Runyon,* 1984]. Copy testing requires substantial funds for allocation to determine consumer reactions to proposed advertisement campaign. It is in this element that copy testing is an important aspect of advertisement management. Yet, it has been a perennial problem in advertising [*Runyon,* 1984).

According to *Aaker, Batra,* and *Myers* [1992], several major questions need to be answered prior to any final advertisement action: Will a proposed copy theme be

effective at achieving advertising objectives? Does the set of advertisements that makes up an advertising campaign create the desired interest level and image? Will an individual advertisement attract the attention of the audience? In simple statement, will the advertisement become effective?

Answer to these questions need to be obtained through the process of copy testing. Thus, copy testing could spell the success or failure of any advertising campaign. It is in this cruciality where significant attention should be provided to advertising copy-testing to ensure that advertising objectives are *realized.* Interest lies in the objective assessment of the validity and reliability of various types of copy tests. Considerable money and time costs are involved in copy testing. Thus, it is imperative to analyze whether to test or not to test, what and when to test and what criteria or test to use.

Much of advertising investment decisions come from results of advertising copy testing. This paper is hoped to provide some insights regarding problems mainly in the validity of print advertising copy testing. With this purpose in mind, advertisers could be given information on how best to undertake probable consumer acceptance and analyze feedback, thus, the appropriate advertising campaign be done. In this way, corresponding financial and time requirement will be determined. More specifically, the purposes of this paper are: (1) to discuss the theoretical consideration or conceptual issues in copy testing, (2) to discuss some experimental results or findings of copy testing, and (3) to develop an alternative conceptual framework of print advertising copy testing that companies may use to improve validity of copy testing. While reliability of copy testing should be considered in the evaluation of print copy media advertising effectiveness, this element may be subjected for further research as the current paper is not involved with it.

The Theoretical Consideration

Past studies have been concerned mainly with the validity of copy testing, however, results have been varied and inconsistent. So far, copy testing is beset with

problems of definition, measurement, and interpretation. Perhaps the most difficult problem in copy testing is determining the objectives against which advertising should be measured. While the ultimate goal of advertising may be to increase sales or market share, advertising may contribute to this goal in a variety of ways. There are number of reasons why sales and market share are misleading criteria of effectiveness of a particular advertisement or advertising campaign [Runyun, 1984]:

(1) advertising does not work alone,
(2) sales response does not always parallel advertising expenditure,
(3) advertising may be subject to threshold effects,
(4) advertisers may find it difficult to associate changes with a specific medium, and
(5) uncontrollable variables may obscure the relationship between advertising and sales.

According to *Aaker* and Myers [1975, as cited by *Lilien*, *Kotler*, and *Moorthy*, 1992], the choice of an ad message consideration should be given to the credibility, the attractiveness, and the power of the source. Then the creative process involves finding the facts and ideas that match the brand's message with its copy objectives. The basis of copy testing is to determine if an ad is likely to work. *Lilien*, *Kotler*, and *Moorthy* [1992] noted that there are two elements involved in copy testing: the dependent-variable measure and the measurement setting. They added that possible measures of response include the following:

(1) attention and impression, the ability of the ad to attract attention and be memorable,
(2) *communication/understanding,* the ability of the ad to convey the message clearly and unambiguously to the target market,
(3) persuasion, the ability of the ad to modify attitudes and beliefs about the product on certain key attributes or to change overall purchase intentions, and

(4) purchase, the ability of the ad to generate a positive impact on purchasing behaviour. The last two measures, while most appropriate, are also the most difficult to measure.

Several studies have attempted to relate desirable characteristics of ads to quantifiable mechanical and massage elements. Most of these studies have focused on readership or recall scores for print ads, the easiest types of ads and response variables to measure. In a study of the effectiveness of industrial print ads, *Hanssens and Weitz* [1980, cited by *Lilien, Kotler,* and *Moorthy,* 1992] used a model ofthe form:

$$Yj = e^{a} \prod_{j-1}^{pt} x_{ij}^{bj} \prod_{j_{Pt.+1}}^{P} (1 + X_{tj}^{1})^{bj'} e^{\mu t}$$

y_i = effectiveness measure for *i*th ad
X_{ij} = value of *j*th nonbinary characteristic of *i*th ad (page number, ad size), $j = 1, \ldots, P_t$.
x_{ij}' = value (0 or 1) of *j'* th binary characteristic of *i*th ad (bleed, color. etc.) $j = P_{t+1}, , P$
e^{a} = scale factor
μ_i = error term

Furthermore, *Diamond* [1968, cited by *Lilien* and *Kotler*] performed a well known study of the effect of advertising-format variables on readership scores. He used data of 1070 advertisements that appeared in *Life* between February 7 and July 31, 1964. For each advertisement he had 6 different Starch readership scores: men-readers noted, seenassociated and read most and women-readers noted, seen-associated and read most. In addition to these 6 Starch scores, he measured 12 variables related to each ad: product class, past advertising expenditure, number of ads in issue, size, number of colors, bleed/no bleed, left or right page, position in magazine, layout, number of words, brand prominence, and headline prominence.

Diamond fitted several regression models and used the coefficient to draw conclusion about the effect of different variables on readership score. Each of the readership models was of the form:

$$R_n = h + \sum_i a_i x_i + \sum_j \sum_k d_{jk} Y_{jk}$$

Where,

R_n = one of the set of 6 readership measures (percent)
h = constant term
x_i = level of ith continuous variable
a_i = contribution of ith continuous variable
d_{jk} = level of jth discrete variable, where k is either 0 or 1, representing the two possible levels
Y_{jk} = contribution of kth state of jth discrete variable.

According to *Aaker* and Day [1991], the copy test validity refers to its ability to predict advertising response. They overview some of the important ways in which copy tests can differ. Each dimension involves validity issues and trade-offs with cost. Several important considerations are involved in evaluating validity such as advertisement used, exposure, sample, measuring response, and appropriateness of the response measure.

Advertisement used—one issue is whether a rough mock-up of an ad or a finished ad is used. The seriousness of the problem will depend on the difference between the mock-up and the finished commercial and the impact of this difference on audience response. Another issue is the frequency of response. To what extent can a copy test predict the response to a campaign that will involve dozens or even hundreds of exposures?. Still another issue is the context in which the test advertisement is embedded. The use of a clutter of advertisements embedded in a

program or magazine is the most realistic but adds complexity and is possibly confounding.

Exposure—*the* approaches such as the theatre tests or mall intercept exposure contexts are termed forced exposure tests because the setting is artificial and the respondent is required to watch. Respondents thus tend to pay more attention to the advertisement than under normal conditions where they *are* much more likely to ignore it completely. Thus, there is still concern that the exposure context may affect the results.

Sample-the sample should be representative of the target population. In all copy test approaches, the biggest concern is with the bias introduced by nonresponse. The danger is that those who refuse to participate may response differently than those who do agree to respond.

Measuring response—there is a concern with the reactive effect of the study. The respondent's reaction may be affected by knowing that actions and opinions are being recorded. Thus, is the respondent willing and able to respond accurately? *Appropriateness of the response measure—* obviously, copy test validity will depend on the advertising response that is desired. A campaign that is designed to gain awareness may not be best measured by a test that focuses on immediate behavioral response. A campaign that attempts to create an image or an association with a feeling such as warmth might require many repetitions and a subtle measurement method. Thus, the usefulness of the various criteria used in testing need to be evaluated in the context of the advertising objectives involved.

Review of Previous Studies

Chambee, Gilmore, Thomas, and *Soldow* [1993] found that the complexity of advertising copy is significantly related to advertising readership. As the ratio of separate words to total words increases, advertising readership increases. This finding should prove useful to advertisers, particularly in light of the fact that the ratio is so easily measured for any ad's copy. What they do not have, at present, is a conclusive

evidence regarding the reasons behind this finding. Their results were, as they expected, based on information theory. However, variation across magazines and gender suggest that there could be interactions between the ratio measurement applied and reader demographics, product category, or involvement level. Finally, they noted that these effects should be explored in future research.

Chow, *Rose*, and *Clarke 11992]* designed the method called Structural Equations Estimation of New Copy Effectiveness (SEQUENCE) to help management evaluate new copy alternatives before placement. The system is intended to: (1) rapidly predict a new copy's effect on brand beliefs, attitudes, and purchase intentions; (2) produce actionable diagnostic information that can be used to improve the copy execution; and (3) permit evaluation of alternative copy executions. They believe the procedure provides more information with respect to the evaluation of advertising copy than traditional approaches at about the same cost; therefore, its diagnostic value is high. The SEQUENCE method was illustrated via case analysis using data from the toothpaste and sparkling-water markets. While using real data may have complicated the explication of the method, they believe that benefits to be gained from examining an application that is filled with the richness and ambiguity of most attempts to assess the effectiveness of ad copy more than compensate the *reader* for the added effort. The application of SEQUENCE to the sparkling-water case data revealed some potentially negative consequences associated with the use of either execution that might have been overlooked in traditional copy-testing procedures.

Fenwick and *Rice [1991]* noted that previous research studies on continuous measurement methods have not presented large-scale test of reliability. Accordingly, advertising practitioners have had to rely, to large extent, on blind faith in accuracy and reliability of this system. They addressed this issue by presenting a large-scale test of 1,164 advertisement evaluations from a continuous measurement system. A continuous measurement copy-testing methodology, the Program Evaluation Analysis Computer (PEAC) system, was discussed and a large-scale test of reliability was conduced Continuous measurement was fount to exhibit quite high levels of test-

retest reliability. That is, reactions to test ads were similar when the ads were tested with matched sample of respondents. Additionally, no sequence effects were found to exist. These results are very encouraging for advertising practitioners who are increasingly using continuous measures in their copy testing. Continuous copy-testing systems present several advantages over the traditional methods of recall and persuasion. The method allows detailed diagnostics and a thorough examination of emotional response throughout the commercial. The use of these methodologies should improve their understanding of emotional response and allow for better determination of effective commercials.

The objectives of the research conducted by Ha/ey, and *Baldlnger* (1991) are to answer the following questions: (1) how well do copy tests, as presently conducted, identify known sales *winner, (2)* which individual measures do the best job, (3) which general types of measures are most predictive, (4) are on-air designs preferable to off-air, (5) are pre/post designs preferable to post only (matched group) designs, (6) are multiple-exposure designs better than single-exposure designs, and (7) is any one copy-testing system superior to the others. According to the authors, 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. They noted that 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. They noted that one thing that most copy-research professional agree upon is that copy testing relates to sales and the ARF study confirms that fact. The study also confirms the PACT Principles endorsed by the research directors of the 21 largest advertising agencies in United States earlier in the 1980s. The principle that states that advertising works on a number of level and, therefore, that no single measure is adequate to measure the effectiveness of copy seems to receive strong confirmation. It was also reassuring to find that, within each of the type 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. The study emphasized that no copy-testing methods nor measures can be rejected as a results of the ARF study. 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. 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.

Haley [1994] stated that if a researcher wants to understand the nature of the copy factors that are driving sales results it is essential that brand effects be removed from the database. He mentions that using comparisons between brand pairs is a good way of accomplishing this. If brand effects are not removed, they are almost certain to contaminate any conclusions reached. Haley also decries misinterpretation of the results, but he is resigned to the fact that people are likely to interpret results in the light of their own experience, predispositions, and biases. However, he attributes the extensive attention that likability is now receiving from practitioners largely to the Bayesian learning that took place as a result of the CVRP. The use of both pragmatic and scientific approaches was intended to call attention to a culture clash between marketing people and statisticians. Most marketing managers, perhaps unfortunately and perhaps not, are not sophisticated in the use of statistical methods. The scientific method, while it sounds good in theory, ends up being somewhat arbitrary. Haley also mentions that their future choices of standards for acceptance and rejection of individual measures lead them into some counterintuitive conclusions. As a final note, he differs sharply with his prediction of a dim future for likability measures. On the contrary, most major copy-testing firms have now added such measures to their standard batteries, and they are joining persuasion and recall as one of the three primary global evaluative measures in use.

Katz, [1991], stated that *"it looks likely that the data-access issue will not go away in the 1990s*". The issue of data access does appear to be one of considerable concerns to those involved in media at major advertising agencies, as indicated both by the 67 percent response rate to the questionnaire and the 65 percent of respondents requesting summary results. Current knowledge of the various problems involved in accessing data were acknowledged by majority of respondents, most notably the differing program configurations and the lack of time to learn new programs. Most media departments conduct their own staff training in the use of computers, and have a computer specialist on their staff as well. Respondents appear interested in gaining both greater amounts of data, such as continuous BAR monitoring and local cable ratings, and even more so, improved access to exiting data, such as BAR on micro and cable data by standard NAD breaks. The issue of the costs involved in handling data was brought out in respondents' comments. Several respondents noted how difficult to monitor all the information being produced, and finding is hard to keep up with all the data and system available. Perhaps the ARF could produce a database (text-retrieval system) with all available research described on it - report by report, product by product. When all the numbers are hooked up, the key factor to remember is why (marketing and media data are collected and accessed in the first place. For on their own, they are not all that important. What matters is the information and knowledge that can be gained from them. Computers are certainly exciting and powerful tools. Let's make sure we make the best use of them to access our data.

According to *Rossiter,* and *Eagleson 11994]* the CRVP provided valuable evidence on the predictive validity of many widely used copy-testing measures. This evidence is mostly negative: it indicates which measures to reject. Using their 90/ 50/10 rule, at least 26 ofthe 35 single-item measures performed so poorly in the CRVP that those can safely be rejected as candidates for *the all-purpose set* of pre-test measures to use in the future. Nine measures, and at least five 2-item multiplicative measures, emerged as worth testing further. About half of these measures worth pursuing involved ad likability in some form or other, and half did

not. It is therefore not justifiable to conclude, as many have, that Liking was the single best measure in the CRVP. The main problem with the small number of ad-pairs tested in the CRVP (only five) is that it leaves *acceptance* of these measures as a risky proposition. How can managers use measures from the CRVP, as analyzed in this article, correctly? First, they would recommend including all *nine* of the copy-test measures. Use of all the promising measures allows the manager to learn from the copy test, by a statistical technique such as regression, how the ad appears to be working (or not working). Second, an in addition to the empirical approach, they would recommend that managers attempt to specify beforehand which of the measures are theoretically essential for the new ad to perform well on. The Liking and Pre/Post Persuasion measures provide a good illustration of this. Across TV commercials in general, Liking and Pre/Post Persuasion exhibit only a modest correlation, the order to 0.2 to 0.4. So, contrary to the results in the CRVP's sample of ads, most ads do not perform well on both measures. The manager has to specify whether both are essential or whether one is essential while the other is not. Managers should applaud the CRVP but be cautious about its findings. Thus, managers cannot safely accept any of the measures in the CRVP based on the CRVP's results alone.

From their analysis, *Walker and Gonten [1989]* have shown that brand-name registration can be iniluenced by the way the name is used in the commercial and, in some cases, by the name itself Further, they have identified a number of executional characteristics that can have a substantial influence on the commercial's ability to attract attention and be remembered in the first place. Even these general findings on the contributors to recall, or its constituent components, should enhance the advertiser's ability to generate higher levels of recall; and the Attention/Linkage Analysis recall model, in itself, is a powerful diagnostic tool with which to explain recall results for individual test commercial. Exception to the overall findings was a small minority of the cases analyzed, but in absolute terms, they represent a substantial number of commercials. They view these exceptions as the most potentially productive ground for further analysis. Their intention is to develop and

improve on the initial findings and test them more rigorously over the complete data distribution, as additional cases are acquired. They have in place a methodology to confirm and explain more about related recall, and its causes, on a continuing basis.

Gronhaug, Kvitastein, and Gronmo [1991] examined how a variety offactors, such as size of advertisement, the number of colors used, the product advertised, and characteristics of the audience, may influence recognition of advertisements, and to explore whether advertisement recognition may be influenced by environmental changes. They found that (1) advertisement readership is positively related to the size of an advertisement, (2) the use of multiple colors is positively related to the reading of advertisements, (3) a positive correlation/beta coefficient is found between women and reading of advertisements, (4) the correlation and beta coefficient between the various product categories and advertising vary as hypothesized, (5) a weak negative but significant coefficient is observed for time and advertising readership, which may be considered as being in concordance with the stated hypothesis. The observation that advertising effectiveness may be influenced by external factors implies that return on investment of a specific advertisement for a given product may change over time. The intuitive implication of this observation is that the relative effectiveness of different advertising vehicles may change over time and should be considered carefully by marketers.

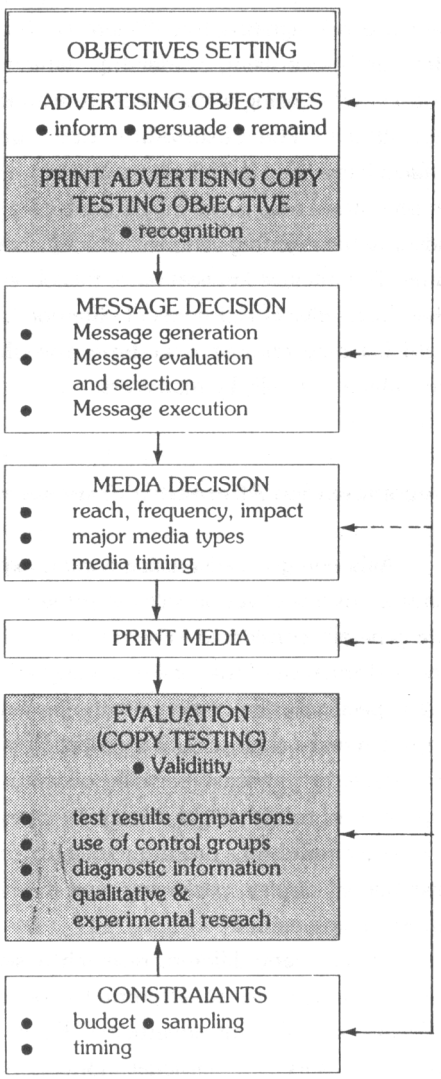
Recommended Alternative Framework

Although good advertising copy is expected to have a positive effect on sales at some point, sales are generally considered to be primarily criterion of the performance of the total marketing effort rather than specifically the copy content. The advertising copy can make an important contribution but usually its effects are at least partially obscured by the influences of media weight, the media types chosen, media scheduling, pricing, sales promotion, distribution, display, packaging, and the combined efforts of competitors.

So far, many different research designs have been employed in an attempt to measure the effectiveness of print advertising. Each of these methods has its own individual shortcomings.

In addition to the presentation of some problems in copy testing validity, the following is an alternative conceptual framework of print advertising copy testing, which includes the planning, execution, and evaluation process of print advertising copy in Figure 1.

Figure 1
Alternative Conceptual Framework for Print Advertising Copy Testing



- ***Objectives Setting***

The first concern in assessing copy-test validity is the objective setting. It should be tested with respect to a communication objective and a copy test evaluated in that context. Thus there must exist an operational objective which is a measurable and useful variable that represents the objective. Clarity in what is being sought from the ad and therefore being tested in the copy test is crucial. One measure cannot generally substitute for another: recall does not measure persuasion, and one must decide which is being sought for a particular ad.

Obviously, therefore, the validity of a particular copy-test will depend on the advertising response that is desired based on the specific objective set. A campaign that is designed to gain awareness may not best be measured by a test that focuses on immediate behavioral response. Recognition refers to whether a respondent can recognize an advertisement as one he or she has seen before *[Aaker, Batra* and Myers, 1992]. Hence, particular attention to recognition as a specific objective of a print advertising copy testing should be emphasized.

- ***Evaluation***

Once the print advertising copy testing objective specifically on recognition has been set, the message has been generated and evaluated with print media decided, evaluation mechanisms should be installed to assess the general results. These mechanisms can come in the form of comparing test results with norms and standards, the use of control groups for experimental research and diagnostic information of response monitoring through computer and other technical programs and processes. Qualitative research through direct and indirect research methodology could be helpful. This is the centerpiece of assessing the validity of print advertising copy.

- **Constraints**

The following could be the identified constraints in the process:

Budget constrains—in applying the scientific method to copy-test, one frequent issue is the cost of data or information. It is obvious that data will be required for analysis. Like every research process, copy-test is an activity which requires comparison of relevant cost with relevant benefit. In copy test (and, for that matter, in many other business research applications) the benefit flowing from the copy testing process is called the value of data. Thus, copy-test validity relates to the money that companies spent to generate data or information. The more money a company spends, the more data or information is affordable on research.

Sampling—sampling considerations include: (1) the definition of the universe from which the sample will be drawn, (2) the method of selection, (3) the size, (4) dispersion (number of location), and (s) location selection criteria.

The definition of the universe from which copy test respondents should be drawn is often a controversial issue. The controversy revolves around whether people in the market target can be expected to respond in a similar fashion to those who are not in the target — especially in sofar as rank order responses to alternative executions is concerned. A'V-?;v qq ^{f}l*'t

Respondent selection methods depend upon the research design and exposure system being used. As the PACT document specifies, however, basic principles of good sampling practice should be observed wherever it is possible and economically feasible to do so. In other word, the sample should be selected so that its results are projectable back to the universe from which it is drawn.

Sample sizes in copy testing have been under cost pressure, especially because doubts have been expressed about the relationships between copy testing scores and sales performance. If the emphasis of the copy tests is on diagnostic use, small samples can be very helpful. According to *Haley* [1985] some agencies routinely use samples of 30 to 50 respondents for all of their copy testing. Sample sizes tend to be larger for captive audience situations involving group exposure,

ranging from 300 upward. The historical standard for on-air tests was 200. In general, however, the trend seems to be toward smaller sample sizes.

The decision about the number of testing locations to be used is another tradeoff situation. The more locations used the more projectable the results are likely to be; however, the more cities the higher the cost is likely to be. Testing in a single location is dangerous because in the absence of contradictory data there is a tendency to assume that results from other cities would be the same if they were available. Furthermore, if results are stable from city to city, we can assume that city difference are not a serious concern in interpreting the results ofthe test. On the other hand, sharply differing results should be taken as a warning that they may not be reliable and that, if the decision is an important one, it will be wise to replicate the test in additional cities.

In selection criteria, where the test is to be conducted in just two markets, preferred practice is to choose market that are about average in term of both brand development and category development. Where more markets are covered, more diversified development situations can be represented. In other words, selection criteria can invoke any factors that experience has shown to be related to responsiveness.

Timing of the Measurements—the timing decisions invoke choices of the interval between exposure and the time at which measurements are taken. For designs that call for *"pre "* measures to be taken, there is question as to how far in advance of exposure they should be taken. And for designs calling for multiple exposures there are questions both as to the length of time that should elapse between exposures and whether or not measures should be taken during that period. These choices revolve around the twin issues of decay and sensitization. Most advertising people believe that the effects of exposure to a commercial decrease as time passes and the exposure is not repeated. This deterioration is called decay. However, although the presence of decay is generally accepted, there is very little research on the results of measuring the effects of exposure to a commercial after the passage of varying amounts of time.

Conclusion

Practitioners are concerned with the measurement of the quality of the message and the copy making up the advertisement. Various measures have been developed to evaluate the impact of an advertisement or advertising campaign on recognition, recall, persuasion, and p'urchase behavior or sales. The sales impact is the most desirable effect to measure, but most of successful work has measured the market's awareness or recognition of the advertisement.

Based on the belief that copy testing validity can be improved, this paper has attempted to outline an alternative conceptual framework, and to draw some variables which may be considered in assessing the effectiveness of a print advertising copy. To improve the validity of print copy testing, the objective, the evaluative measurement and the constraints have to be considered as these influence to a great extent the validity of the print advertising copy . It, thus, achieve the objective of recognition. Recognition is a necessary condition for effective advertising. According to *Aaker, Batra* and Myers [1992], if the advertisement cannot pass this minimal test, it probably will not be effective. This can only be ascertained through the validity of print advertising copy testing.

SELECTED REFERENCES

Aaker, David A; Batra, Rajeev; and Myers, John G [1992]. **Advertising Management,** Fourth Edition, pp. 406-438.

Aaker and Day [1991]. **Marketing Research** *Advertising Research, Copy Test Validity,* pp. 709-735.

Chambee, Robert; Gilmore, Robert; "Thomas, Gloria; and Soldow Gary [1993]". *When* Copy *Complexity Can Help AD Readership,* **Journal of Advertising Research,** Vol. 33, No.3 (May/June 1993),pp. 23-28.

Chow, Simon; Rose, Randall L; and Clarke, Darral G [1992]. *"Sequence: Structural Equation Estimation of New Copy Effectiveness",* **Journal of Advertising Research,** Vol. 32, No.4 (July/August 1992), pp. 60-72.

Fenwick, lan and Rice, Marshall D [1991] . *"Reliability of Continuous Measurement* Copy Testing *Methods",* **Journal of Advertising Research,** Vol. 31, No. 1 (Feb/March 1991), pp.23-29.

Gronhaug, Kjell; Kvitastein, Olav; and Gronmo, Sigmund [1991]. "Factors Moderating Advertising *Effectiveness As Reflected in 333 Tested Advertisements".* **Journal of Advertising Research,** Vol. 31, No. 5 (October/ November, 1991), pp. 42-50.

Haley, Russell I [1985], **Developing Effectiveness Communications Strategy A Benefit Segmentation Approach.** Ronald Series on MARKETING MANAGEMENT, John Willy & Sons,, pp. 279-353.

Haley, Russell I and Baldinger, Allan L [1991], *"The ARF Copy Research Validity Project",* **Journal of Advertising Research,** Vol. 31, No. 2 (April/May 1991), pp. 11-32.

Haley, Russell I [1994], *A Rejoinder to Conclusions From the ARF's Copy Research Validity Project,* **Journal of Advertising Research,** Vol. 34, No.3 (May/June 1994), pp. 33-34.

Honomichl, Jack J [1986], **Honomichl on Marketing Research,** pp. 267 280.

Katz, Helen [1991], "Hoiu *Major U.S. Advertising Agency Are Copying With Data Overload",* **Journal of Advertising Research,** Vol. 31, No. 1 (February/March 1991).

Kinnear, Thomas C and Taylor James R [1991], **Marketing Research,** "An *Applied Approach",* pp. 765-777.

Kotler, Philip [1994] **Marketing Management Analysis, Planning, Implementation and control.** 8'h Edition, pp. 626-652.

Lilien, Gary L; Kotler, Philip; and Moorthy, K. Sridhar [1992], **Marketing Models,** chapter 6,pp. 295-306.

Rossiter, John R and Eagleson, Geoff [1994], *"Conclusions From the ARF's Copy Research Validity Project",* **Journal of Advertising Research,** Vol. 33, No. 3 (May/June 1994), pp. 19-32.

Runyon, Kenneth E [1984], **Advertising,** Second Edition, pp. 592-612.

Walker, David and Gonten, Michael.F.von [1989]. *Explaining Related Recall Outcomes: New Answers From A Better Model.* **Journal of Advertising Research,** Vol. 29, No. 4 (June/ July, 1989), pp. 11 -21.