

THE MEASUREMENT OF ADVERTISING EFFECTIVENESS

Dr. CHRISTOPHER ORPEN
University of Cape Town

A number of different procedures have been developed to test the relative effectiveness of different advertisements. Unfortunately, despite recent improvements and so-called refinements, all suffer from serious drawbacks. It is my contention that these drawbacks are sufficiently serious for clients *not* to accept any one of these techniques alone as accurate indications of advertising effectiveness.

Considering each procedure in turn:

- (a) Client jury technique — the personal views of the client who commissioned the ads is sought as to which of a number of alternative ads he prefers or likes most and least.
- (b) Consumer jury technique — the personal views of a small but fairly representative sample of potential buyers of the advertised product are solicited as to which of a number of alternative ads they prefer or like most and least.
- (c) Readership surveys — for ads which appear in newspapers or magazines. Readers of the given newspapers or magazines are asked whether they can recall or recognize the various ads.
 - (i) In aided recall tests (Gallup) each subject who reads a given issue is provided with a set of brand or company names and asked whether he remembers seeing an ad for any one of them in the issue. For each ad that S claims to have seen, he is required to 'play back' its selling message from memory.
 - (ii) In recognition tests (Starch) each S who has read a given issue is shown each ad in turn and asked whether he remembers seeing it, whether he can name the product or advertiser and how much of the copy he read.
- (d) Controlled exposure studies — a group of Ss are presented with the alternate ads under 'experimental' conditions, usually for only a brief period by means of a tachistoscope. On the basis of this very short exposure, they are asked what

aspects of the ad they can remember and how much they like or dislike each.

(e) Buying behaviour studies:

- (i) brand-use survey — consumers are asked, usually at home, what brands of various products they are currently using and their responses compared with the degree to which various newspapers and magazines featuring ads for the brand are bought and read. Generally, if brand A is advertised in newspaper X the frequency of use of A among readers and non-readers of X is compared.
- (ii) Coupon returns — the various ads are designed to carry coupons, which have to be cut out and returned to the advertiser for 'free samples', 'further information', etc. Generally, a comparison is made of the number of coupons returned for the various ads.
- (iii) Sales tests — a comparison is made of relative sales of the brand in similar market areas in which ads for the brand have and have not appeared in local magazines and newspapers.

In order to evaluate the extent to which each of these standard techniques can yield accurate results, it is necessary to set down what empirical research has revealed about how advertising works.

Important points to note:

- (a) Advertisements are effective insofar as they move people closer to actually buying the advertised product. Generally, an ad is effective if it induces Ss to go out and buy the product and ineffective if it fails to induce them to do so. This sounds obvious, but it is often overlooked. The ultimate aim of an ad is *not* to produce something that is necessarily liked, admired, recognized, enjoyable, eye-catching, or even immediately meaningful — it is to sell a production and the achievement of these above 'qualities' is only

important insofar as a definite relationship has been established, through empirical research, between these so-called 'qualities' and inducing people to go and actually buy the product. As will be shown later, few such definite relationships have been established.

(b) Consumers tend to expose themselves to ads in which they are initially interested or which they find congenial to their existing attitudes and to avoid ads in which they are initially disinterested or find uncongenial to their existing attitudes. For instance, Ehrich (1957) found that new car owners were more likely to read ads for the car they had just purchased than were owners of the same make but an earlier model. The new car owners were also much more likely to read ads about their own car than they were about other makes. They argued that this was largely because the new car owners were seeking reassurance by exposing themselves to what were, for them, very congenial communications. This general tendency for consumers to only expose themselves to messages which confirm their attitudinal predispositions has been confirmed in a variety of 'advertising' areas; political broadcasts (Lazarsfeld, Berelson and Gander, 1949), newspaper articles (Cannell and MacDonald, 1956), government advertisements (Cartwright, 1949) and educational campaigns (Star and Hughes, 1955).

(c) Even if the consumer is somehow 'forced' to expose himself to given ads, he is likely to misperceive or reinterpret their messages to make it congenial with his existing attitudinal predispositions. This tendency to selectively perceive ads in this way is neatly illustrated in an experiment by Horland and Janis (1958) in which ads or communications arguing the desirability of prohibition (abstinence) were presented to three types of persons; heavy drinkers, slight drinkers, and abstainers. They found that the greater the difference between the attitude of the recipient and the position advocated by the communication or ad, the more likely the recipient was to regard it as being propagandistic and unfair; and even to perceive the stand advocated by the ad or communication as further removed from his own position than it actually was. Conversely, when the distance was small between the recipient's own position and the stand advocated by the communication, the recipient was likely to view the ad or communication as being fair and factual and to perceive it as being closer to his own position than it actually was. This general result has also emerged from a number of other studies, notably those by Carlson (1956), Jones and Kohler (1958), Vroom (1960) and Edwards (1941).

(d) There is another way in which a consumer can

reduce the discrepancy between his own attitudes and those expressed by the ad or communication and that is simply its content rather quickly. If this is the case, then we would expect a person to learn more quickly and remember for a longer period ads or communications which are comparable with their own attitudes. This is exactly what was found in a classic study by Levine and Murphy (1943). In their study pro-communist material was better learned and better remembered by persons with pro-communist views than by those with anti-communist views, while the opposite was true of anti-communist material. Among other studies to confirm the selective retention of congenial and uncongenial material in this general way are those by Garber (1955), Bauer (1958) and Zimmerman and Bauer (1956).

(e) All the above psychological research has been repeated by more recent experimenters, with the same general result, namely that consumers are *very* capable of resisting attempts to change their attitudes and behaviour. This research suggests that a great deal of advertising only serves to *reinforce* existing attitudes and behaviours (e.g. maintain brand loyalty) or to *stimulate* or activate consumers who are already predisposed to act in the desired manner anyway (e.g. people who like bright colours to be influenced by ads of brightly-coloured cars). A related implication is that advertising is not, in itself, a cause of consumer effects but works with, and through, various mediating factors such as consumer predispositions and personal influence, like word-of-mouth advertising.

(f) Even in the case of reinforcing existing attitudes or behaviour which (at least) appears to be within the scope of advertising, research has indicated that whether an individual expresses a liking or disliking for a given ad, especially shown out of its 'natural' environment, is not closely associated with attitudes more closely related to the ad's effectiveness, such as expressed desire for the product after real-life exposure to the ad (Treasure and Joyce, 1967; Orpen, 1974). This is mainly because, as many studies have consistently shown, the decision to buy or not to buy a given product (which decision is the focus of advertising) is determined, to a large extent, by deep-lying motivational factors, which are frequently unrelated to verbal statements about liking or disliking that may be made to a particular investigator (Martineau, 1957; Ferber and Wales, 1958; Henry, 1958; Smith, 1954; Haire, 1950). This discrepancy between expressions of liking on the one hand and deep-lying preferences, on the other, is made worse by the fact that, in a testing situation, consumers often tend to give answers that they think will 'please' the examiner rather than those that ac-

curately reflect their own views (if they know them). This discrepancy, usually investigated under the heading of 'social desirability' has been revealed in a number of studies (e.g. Cheskin and Ward, 1948; Langstaff and Layhorne, 1949; Edwards, 1957, 1959; Orpen, 1971a, 1971b).

(g) In principle it should be noted that it certainly does not follow that an ad which is effective in persuading or influencing consumers to buy a given brand need be remembered better than one which is not. It should also be remembered that the point is not that the advertisement *as such* should be remembered, but rather the product which is featured. Bearing these two points in mind, plus the fact that the variables determining the crucial decision (for advertisers) of whether to buy a given brand or not, are very different from those determining the trivial decision (for advertisers) of whether to admit to recognizing or recalling a given advertisement or not in a testing situation, it is not surprising that Haskins (1964) should conclude, on the basis of a number of studies with advertising and other material, that 'learning and recall of factual information from advertisements occurs, but recall and retention measures seem, at best, irrelevant to the ultimate effects desired, the changing of attitudes and behaviour' (p.6). Haskins claims that this is largely because the kind of 'learning' involved in advertising is of the incidental kind rather than of the highly-motivated kind of learning that takes place in, say, a classroom situation (in which recall and retention measures *are* related to effective learning). That recall and recognition tests are irrelevant also stems from the fact that consumers do not direct their attention to specific material in ads (which generally occurs in non-incidental learning), but instead gain a vague impression of the ad, with it either impinging or failing to impinge on their consciousness as they read the newspaper, listen to the radio, attend a movie or drive a car. As Joyce (1967) states 'in general there seems sufficient evidence available to lead one *not* to rely upon recall measures as indicators of effectiveness' (p.174).

Let us now examine each of the main techniques in the light of this evidence.

(a) Both the client and consumer jury techniques and the readership surveys are basically tests of liking and preference, generally for advertisements as such. Hence, they cannot, in their very nature, give an indication of advertising effectiveness. In this sense, they are a waste of time and, if their results are taken seriously, can do more harm than good.

Moreover, the jury techniques especially suffer from the very serious disadvantage of being

contrived or artificial, in the sense that the ads are seen out of their 'natural context'. In this respect, it should be noted that there is a world of difference between the laboratory situation, in which ads are presented *on their own* to judges and the real-life situation, in which the ads are seen amongst other competing material on the pages of a newspaper and magazine or on billboards. In addition to the different contexts, the attitude of the jury member and the reading public are vastly different. The former are carefully evaluating the ad from some personal standpoint (as critics), the latter are quickly scanning pages and billboards with their attention usually focussed elsewhere (as spectators). It is not surprising, therefore, that the correlation between rated liking for given ads (consumer jury) and the sum of expressed willingness to buy the advertised product and number of coupon returns for the same ads was virtually negligible (Orpen, 1974).

Other disadvantages concern the extreme subjectivity of the judgments and the fact that, in the case of the client jury, the person who commissioned the ad is frequently unable to accurately 'put himself' in the role of the potential customer whereas, in the case of the consumer jury, the members of the group (usually pretty small) are often not fully representative of the potential buyers of the product. Finally, it is not possible to discount that the judges may be answering in such a way as to 'please' their bosses or enhance their standing in the firm (client jury) or make a good impression on the investigator (consumer jury).

(b) Both kinds of readership surveys rely on recall and recognition tests, usually of the advertisement itself. Hence, in view of the research mentioned earlier they cannot, in their very nature, be accurate indicators of advertising effectiveness. In addition, readers frequently report having seen an ad that did not appear at all, in an attempt to impress the investigator (Lucas, 1940; Lucas and Murphy, 1939). This 'desirability' effect is more pronounced in the case of ads for well-known brands than for ads for more unfamiliar brands (Kopoven, 1956).

(c) What is needed is a technique that focusses more squarely on the ultimate aim of advertising — the actual buying of brands and products. As Anastasi (1961) says 'The ultimate question is whether the ad succeeds in leading to appropriate action. Although more difficult to measure under controlled conditions than the mere attention or memory value of ads, buying behaviour provides the most comprehensive index of and effectiveness' (p.257).

The trouble is that the current behavioural techniques — brand use surveys, coupon returns,

and sales tests — are each fraught with difficulties. For instance, brand use surveys are open to difficulties in interpretation, quite apart from the fact that consumers do not always answer honestly; e.g., the fact that a high correlation is found between the use of a given brand and reading newspapers or magazines containing ads for the brand does not necessarily imply that it was the ad that induced the purchase. It may have well been due to a host of other factors, other than seeing the ad in the newspaper or magazine. Coupon responses are often inaccurate as *sole* indicators of the effectiveness of a given ad because the coupons are frequently returned by subjects who are not likely to buy the product as a result of seeing the ad; e.g., coupons are often returned by children as a 'game', by Africans who enjoy the prestige of getting mail and habitual coupon-clippers. Also, frequently the nature of the offer and the physical characteristics of the actual coupon are not held constant across the various ads and hence valid comparisons cannot be made. Finally, sales tests are very difficult to carry out and, in any case, are not of such a nature that the effects of extraneous factors like changes in the weather, alterations in fashions, variations in the general economic climate on sales (rather than the advertisement *per se*) cannot be ruled out.

In the light of these difficulties and the research findings presented earlier, the following guidelines for the accurate measurement of advertising effectiveness can be laid down.

- (1) The artificial and unnatural situations used in the client and consumer jury techniques should be avoided.
- (2) Judges should preferably *not* be asked to evaluate the merits of the advertisement as such or to express the extent of their liking or disliking for it.
- (3) More importantly, such evaluations and expressions should *not* be used to decide which ad is likely to be more effective than another.
- (4) Recall and recognition tests — readership surveys and controlled exposure studies should also *not* be thought of as giving an accurate index of effectiveness.
- (5) The techniques should focus on the decision to buy or, more specifically, on whether given ad is likely to induce a person to buy the advertised product.
- (6) The technique should enable more control to be exercised over the initial factors than is possible with the current brand use surveys and sales tests.
- (7) Given the limitations of each method, the researcher should employ *more* than one method and should not accept any results as accurate unless they emerge from all (or both) techniques he

employees; i.e., he should look for consistent results across his methods.

On the positive side, the following approach is recommended: Assuming that the researcher wants to establish which of two (A or B) ads for a given brand (say, of wine) is more effective, he should:

- (i) Obtain a representative sample of drinkers and non-drinkers of the brand.
- (ii) Randomly divide the sample into six sub-groups, each consisting of equal numbers of drinkers and non-drinkers.
- (iii) Send identical newspapers to sub-groups 1 and 2, the only difference being that the newspaper received by 1 contains ad A and that received by 2 ad B. Both ads contain the *same* coupon in the same position (e.g. bottom right hand corner), making the same request (e.g. send the coupon back for a small sample).
- (iv) Send identical newspapers to sub-groups 3 and 4, the only difference being that the one newspaper contains ad A (without a coupon) and the other ad B (also without a coupon). The next day visit the subjects in both groups personally and simply ask the extent to which they would be willing to buy the advertised wine.
- (v) Visit groups 4 and 5 personally, take them through the above newspaper (featuring ad A for the one group and ad B for the other), and simply ask each subject 'Having seen the various ads in the newspaper, tell me to what extent the ad (A or B) would induce you to buy the advertised wine'.
- (vi) The comparison between the number of coupons returned by groups 1 and 2 would give you an indication of the extent to which each of the two ads, in their natural setting, get attention and induce action.
- (vii) The comparison between the relative degrees to which groups 3 and 4 were willing to buy the advertised wine gives a second indicator of the action potential of the two ads, without the subjects in any way having to evaluate the ad as such.
- (viii) The comparison between the relative degrees to which groups 5 and 6 said that the ad with its attention-getting property (measured in the 1 vs 2 and 3 vs 4 group comparisons) held constant, would induce them to buy the advertised wine gives a third indication of the relative effectiveness of the two ads.
- (ix) Each of these three comparisons, which measure slightly different aspects, should be studied and themselves compared. Only if each of them are in favour of one of the ads rather than the other can it be concluded unequivocally that one ad is definitely more effective than the other.
- (x) This technique- or comparison of three techniques — does not ask the subjects whether they like the ad as a whole or whether they can

remember it, instead it focusses on whether the ad induces action on the part of the reader in three slightly different ways; action which approximates real life buying behaviour more closely than the jury techniques, readership surveys, or controlled exposure studies currently in use and avoids those aspects of current buying behaviour studies, which make accurate comparisons impossible.

REFERENCES

- EHRlich, D. Post decision exposure to relevant information. *Journal of Abnormal and Social Psychology*, 1957, 54, 98—102.
- LAZARSFELD, P.F., BERELCON, B.R., & GANDET, H. Radio and the printed page as factors in political opinion and voting, in Schram, W. (ed.) *Mass Communication*. Urbana, Illinois: University of Illinois Press, 1949, 481—485.
- CANNELL, C.F., & MACDONALD, J.C. The impact of health news on attitudes and behaviour. *Journalism Quarterly*, 1956, 33, 315—323.
- CARTWRIGHT, D. Some principles of mass persuasion: selected findings on the sales of U.S. War Bonds. *Human Relations*, 1949, 2, 253—267.
- STAR, S.A., & HUGHES, H.M. Report of an educational campaign: the Cincinnati plan for the United States. *American Journal of Sociology*, 1950, 55, 389—400.
- HOVLAND, C.I., & JANIS, I. *Personality and persuasibility*. New Haven, Conn.: Yale University Press, 1958.
- CARLSON, E.R. Attitude change through modification of attitude structure. *Journal of Abnormal and Social Psychology*, 1956, 52, 256—261.
- JONES, E.E., & KOHLER, R. The effects of plausibility on the learning of controversial statements. *Journal of Abnormal and Social Psychology*, 1958, 57, 315—320.
- VROOM, V.H. The effects of attitudes on the perception of performance goals. *Human Relations*, 1960, 13, 229—240.
- EDWARDS, A.L. Political frames of reference as a factor influencing recognition. *Journal of Abnormal and Social Psychology*, 1941, 36, 34—50.
- LEVINE, J.M., & MURPHY, G. Learning and forgetting of controversial material. *Journal of Abnormal and Social Psychology*, 1943, 38, 507—517.
- GARBER, R.B. Influence of cognitive and affective factors in learning and retaining attitudinal materials. *Journal of Abnormal and Social Psychology*, 1955, 51, 384—389.
- BAUER, R.A. The communicator and the audience. *Journal of Conflict Resolution*, 1958, 2, 67—77.
- ZIMMERMAN, C., & BAUER, R.F. The effect of an audience upon what is remembered. *Public Opinion Quarterly*, 1956, 20, 328—348.
- TREASURE, J.A. & JOYCE, T. *As others see us*. London: Institute of Practitioners in Advertising, 1967.
- ORPEN, C. Expressed liking as a measure for advertising effectiveness: a psychological analysis. Unpublished research, University of Cape Town, 1974.
- MARTINEAU, P. *Motivation in advertising*. New York: McGraw-Hill, 1957.
- FERBER, R., & WALES, H.D. (eds.) *Motivation and market behaviour*. Homewood, Illinois: Irwin, 1958.
- HENRY, H. *Motivation research*. New York: Ungar, 1958.
- SMITH, G.H. *Motivation research in advertising and marketing*. New York: McGraw-Hill, 1954.
- HAIRE, M. Projective techniques in market research. *Journal of Marketing*, 1950, 14, 649—656.
- CHESKIN, L., & WARD, L.B. Indirect approaches to market reactions. *Harvard Business Review*, 1948, 26, 572—580.
- LONGSTAFF, H.P., & LAYBOURN, G.P. What do readership studies really prove. *Journal of Applied Psychology*, 1949, 33, 585—593.
- EDWARDS, A.L. *The social desirability variable in personality research*. New York: Dryden, 1957.
- ORPEN, C. The fakability of the EPPS in the personnel selection context. *Personnel Psychology*, 1971a, 24, 1—4.
- ORPEN, C. Social desirability in the Psychological Screening Inventory. *Journal of Clinical Psychology*, 1971b, 27, 463—465.
- HASKINS, J.B. Factual recall as a measure of advertising effectiveness. *Journal of Advertising Research*, 1964, 4, 2—8.
- JOYCE, T. What do we know about how advertising works, in Ehrenberg A.C., and Pyall, F.G. (eds.) *Consumer behaviour*. Harmondsworth, Middlesex: Penguin Books, 1971.
- LUCAS, D.B. A technique for measuring the impression values of specific magazine advertisements. *Journal of Applied Psychology*, 1940, 24, 778—790.