SURENDRA N. SINGH and MICHAEL L. ROTHSCHILD*

Recognition has not received extensive acceptance by practitioners as a response variable for learning of television commercials. The authors review some of the distinctions between recall and recognition, and some of the problems associated with each. In a laboratory experiment they developed and tested a measure of recognition useful for low involvement situations in which a recall level of learning may not be necessary. Data and implications about the usefulness and limitations of the measure are presented.

Recognition as a Measure of Learning from Television Commercials

Many billions of dollars are spent each year on advertising. A large percentage of that amount is devoted to the repetition of messages to enhance learning and inhibit forgetting. The currently most common measure of this learning is recall. Zielske (1959) has shown that the level of recall in a target market increases with repetition and decreases in its absence. Over time practitioners have developed a reliance on recall scores (especially day-after recall of television commercials) to show the relative merit of different messages and different media schedules.

In recent years, however, several studies have suggested that recall may not be the most appropriate measure of learning effectiveness (see, for example, Haskins 1964; Krugman 1972) and it has been proposed that recognition may be better in certain situations: ". . . the inability to recall something does not mean it is forgotten or that it has been erased from memory. The acid test of complete forgetting is if you can no longer recognize the object" (Krugman 1972, p. 14). Bettman (1979) also suggests that recognition is more important than has previously been thought.

Bettman has suggested that the way in which information is to be used should dictate whether a recall or

recognition level of learning is necessary. For example, a brand choice decision made in a store may need only a recognition level of learning because the choices may be reviewed at the time of decision-making. Alternatively, at-home decisions may require a recall level of learning because most choices are not physically present in the home for examination. Bettman suggests that for consumer convenience goods, a shopping list of products is typically made in the home but final brand selection generally occurs in the store. If this is true, the lower level of learning which is tapped by recognition may be sufficient. Bettman's distinction between recall and recognition is used in this article.

The difference between recall and recognition is that in a *recall* test a subject is first given a set of information and is later given some minimal cue and asked to retrieve and reconstruct the original information, whereas in a *recognition* test the subject is confronted with the original material and asked whether it has been seen or heard before. Hence, for *recall* the individual must describe the stimulus which is not present; for *recognition* the stimulus must merely be identified as having been previously seen or heard (Bettman 1979).

Though a recall test of learning is the current standard of television advertising effectiveness (print has a history of being measured by recognition), it may be too stringent in that it may mask some components of memory. Trying to maintain learning at a level sufficient to achieve recall scores at high levels therefore may be "overkill" because a recognition level may be sufficient. Sufficiency of a lower level of learning would have great implications for media strategy. The use of recognition may lead to more modest advertising expenditures than are

needed for recall (Krugman 1977a) and may result in:

—reduced media budgets, or
—a reallocation of dollars to broader *reach* objectives because *frequency* levels could be reduced.

The recall versus recognition issue also is related to the concept of involvement. Early marketing/advertising literature reflects an underlying implicit assumption of high involvement on the part of the consumer (see, for example, the extensive hierarchy of effects and multiattribute model literature). Only in recent years has an assumption of low consumer involvement been proposed.

Recall seems to be a more appropriate measure of learning for products which elicit high involvement. If one accepts the idea that many consumers have low involvement for many products, it may follow that consumers would not need a recall level of learning for these products of marginal interest. Recognition may be a more appropriate measure of learning in these cases.

Another cue to the potential value of recognition comes from the literature on hemispheric lateralization in the brain. It has been suggested (see, for example, Krugman 1971, 1979 and Hansen 1981) that television viewing, visual concepts, emotional concepts, and recognition tasks are dominated by the right hemisphere; reading print, rational and cognitive issues, and recall tasks are dominated by the left hemisphere. Krugman has suggested that recall is the more complex task and is more of a left-brain type of activity. Therefore, he continues, measurement of print effects should be done through recall tasks. Recognition is less complex and more of a right-brain activity. Because TV may activate right-brain activity, perhaps it should be measured by a recognition task. Though data to support these suggestions are not currently found in the marketing/advertising literature, and though it is likely that a complex stimulus such as a television commercial elicits both right and left hemispheric activity, the concepts put forth by Krugman lead one to a consideration of recognition as a test for television commercials.

Finally, another piece of evidence suggesting that recognition may be a better test for learning of television commercials comes from the incidental versus intentional learning paradigm (Eagle and Leiter 1964). Incidental learning is better tapped by recognition tasks. Because most of the learning taking place from television commercials is incidental (Beals et al. 1981; Haskins 1964), it seems proper to measure television commercial effectiveness by recognition measures. We explore the development of a useful recognition measure of television commercial effectiveness for products toward which people generally have low involvement.

A REVIEW OF PAST RECOGNITION MEASURES

"Advertising researchers have argued about the relative merits of using recognition or recall as a measure of advertising effect. Because the criterion of recognition is more easily achieved than that of recall, it has been criticized as less sensitive" (Krugman 1977b). In other words, recognition measures may not be able to discriminate across commercial stimuli.

Recognition has been a widely studied phenomenon in psychology and education psychology. Most of the studies have used words, nonsense syllables, and numbers as stimuli (see, for example, Craik 1971; Eagle and Leiter 1964; Shepard and Cheng 1963). Others have used pictorial stimuli (see, for example, Haber 1970; Shepard 1967; Tversky 1973). One basic feature of these studies is that they all have resulted in very high recognition scores (especially those studies using pictorial stimuli) even after long periods of decay time. For example, Shepard (1967) tested retention of picture stimuli over a period of 120 days. He tested groups of subjects with no delay and also at retention intervals of two hours, three days, one week, and 120 days. Recognition immediately after exposure was about 98%. Although forgetting took place, it did so very slowly with 57.7% recognition after 120 days. Similar results have been reported by Haber (1970) and Strong (1912). Haber showed 2560 photo slides at the rate of one every 10 seconds during four-hour viewing sessions held on consecutive days. In a recognition test, one hour after the last slide was shown, subjects could recognize the pictures with 85 to 95% accuracy.

Advertisers therefore have been discouraged in their use of recognition measures because of the indiscriminately high scores reported. It is important to note, though, that these studies had methodological problems such as a very long exposure time allowed to the subjects (Shepard's 1967 study allowed self-paced viewing of stimuli; Haber 1970; Standing, Conezio, and Haber 1970; Dallett, Wilcox, and D'Andrea 1968 used 10 seconds per picture as exposure time); small numbers of subjects used (Shepard 1967 used four subjects per cell; Standing, Conezio, and Haber 1970 used only two subjects per condition); no controls for previous exposures (Shepard 1967; Strong 1912), highly memorable material (Shepard 1967), or prior subject knowledge of the test.

Recognition scores need not always be extremely high in comparison with recall scores. For example, one can use distractors that are very similar to list items; Goldstein and Chance (1970) did this with snowflakes. In the case of word recognition, using distractors that are strongly associated with list items (e.g., "cat" as a list word and "dog" as a distractor) will deflate recognition scores (e.g., Underwood 1965; Underwood and Freund 1968). Finally, using a large number of distractors (for example, presenting an old item with 100 alternatives (Davis, Sutherland, and Judd 1961) will lead to lower recognition scores.

Klatzky (1980) has developed three classifications of recognition testing procedures:

1. *Yes/no recognition test.* Subjects are shown a series of items, one at a time. As each item appears, the subject is to respond "yes" if s/he thinks it was on the original

list and "no" if it was not. Usually, half of the items on the test are from the original list and the other half are distractors.

2. *Forced-choice test*. In this procedure, subjects see two or more items at a time during the test. The task is to pick out the original stimulus. If the subject sees two items at a time, the test is called a two-alternative forced-choice; if three, a three-alternative forced-choice; and so on.

3. *Batch-testing procedure*. In this method, all the original stimuli and all the distractors are presented at once. The subject then tries to indicate which items were on the original list.

A Comparison of Recall and Recognition

Currently, the dominant theoretical model comparing recognition and recall is the dual process hypothesis (Anderson and Bower 1972, 1974; Kintsch 1970, 1974). According to this view, recall consists of search and recognition. The subject, when faced with a problem, generates a number of prospective candidates for recall during the search process and then recognizes one of them. The decision stage of recall is assumed to involve the same processes as are involved in recognition. Hence, in the dual process hypothesis recognition is assumed to be a subprocess of recall. The implications are that if an item can be recalled it must also be recognized and that the total number of recognized items should always be greater than or equal to the total number recalled.

Tulving and Thomson (1973) and Wiseman and Tulving (1976) have provided data that violate both of these implications. Basically, it has been demonstrated that in certain situations an item can be recalled but not recognized. The phenomenon has been called *recognition-failure of recallable words*. Several criticisms of the "recognition-failure" paradigm are related to its experimental procedures (Martin 1975; Reder, Anderson, and Bjork 1974) and some also are related to its theoretical implications (Santa and Lamwers 1974, 1976). According to Santa and Lamwers, in the dual process model recognition is assumed to occur implicitly when an item is recalled. Successful recall then means that the internal recognition was successful. Therefore, a failure to recognize an item (even though it was recalled) in an external test will not necessarily invalidate the dual process model unless the external recognition test precisely simulates the internal recognition process that goes on during the recall process (Klatzky 1980). Anderson and Bower (1974) more recently have presented a modified dual process theory whereby a word may be represented by a number of different nodes in the memory, corresponding to different senses or "ideas" of the word. Occasional recognition failure is assumed to be due to retrieval failure of information available in memory store. The modified dual process theory thus suggests that recognition, too, may involve both search and decision processes and that recall and recognition may involve similar processes (see Atkinson and Juola 1973; Mandler 1972, 1980; Mandler, Pearlstone, and Koopmans 1969).

Mandler (1980) provides a rich review of two-stage theories.

Recognition and the Advertising Industry

Recognition has been used on a fairly limited basis for testing television commercials. Gallup and Robinson use aided recall of television commercials, and in the past several years Bruzzone Research Company (BRC) has begun using recognition as a measure of learning effectiveness. BRC mails questionnaires that have storyboards of commercials on them along with several multiple-choice questions to a nationwide cross-section of households. All identification signs are blocked out of the storyboards and subjects are asked to indicate whether they remember seeing the commercial and whether they can recognize the brand name, which is presented with two distracting brand names.

Recognition has been more widely used in print media by firms such as Starch INRA Hooper. Published ads and miniatures of posters are presented to respondents who are asked whether they recognize the stimuli. If respondents say they have seen the particular advertising message before, they are counted as part of the advertising audience. (Sometimes an additional question is asked to qualify the person as actually recognizing the material and not just saying so.) Seeing and thorough reading of a magazine ad are usually reported as "noting" and "read most."

Lucas and Britt (1963, Ch. 3) and Clancy et al. (1979) discuss in detail the limitations of recognition tests in print media. Lucas and Britt conclude that there is a general inflation in recognition scores attributable to "people who could not possibly have seen particular advertisements," and that the range of "false" recognition may run from five to as much as 50% (Lucas and Britt 1963, p. 58).

The Printed Advertising Rating Methods (PARM) committee of the Advertising Research Foundation (ARF) found that recognition scores for noting of printed ads did not decline as time passed after the reading of the ads (Lucas 1960). On the basis of this evidence, Lucas and Britt claimed that "recognition ratings may remain at about the same level for months; there is not the memory loss that might be expected" (p. 58). The lack of decline, however, may have been due in part to a response bias. The subject may have been biased toward making a particular response, or may have been better at recognizing "old" items than "distractors" (where "old" item refers to the stimulus item and "distractors" refers to the other nonstimulus items included in a recognition test). Appel and Blum (1961) call this response bias in a recognition test a "noting set tendency" or a tendency to note ads. Their data support the idea that certain readers have a higher tendency to note ads, whereas others have a lesser tendency toward ad noting regardless of whether they are actually exposed to the ad. Moreover, in the PARM study, the introduction of a significant percentage of nonreaders of the measured magazine issue

seems to have had no appreciable effect on recognition as long as these individuals "claimed" to be readers.

In summary, the literature suggests two major problems associated with recognition measures:

1. Recognition measures tend to yield indiscriminately high scores.
2. Recognition measures are affected by response bias.

These problems can be overcome in several ways. For example, recognition scores can be lowered by making recognition tests more difficult, either by including a large number of distractor items or by using distractor items that are very similar to the stimulus item. The problem of response bias also may be eliminated by use of a multiple-alternative forced-choice method; here the stimulus item in the recognition test is presented simultaneously with one or more distractor items. Presumably, the overall tendency to classify an item as "old" should affect both alternatives alike and so should exert little influence on observed choice (Shepard and Cheng 1963, p. 93).

OBJECTIVES

Advertisers have not wanted to use recognition scores as measures of memory because the scores tend to cluster about some upper level asymptote (a "ceiling" effect). The first objective of our study, therefore, was *to develop a discriminating recognition measure*, i.e., a measure that can access the memory for a commercial, but is difficult enough not to produce a ceiling effect. As shown in the literature review, this can be done by using a large number of distractors in the multiple-alternative forced-choice recognition test and by testing recognition after a delay (a two-week delay was used in our study). The measure we present was tested for its discriminating ability. One way to test the discrimination ability of such a measure is to test across commercials. However, a better way is to test (1) across a number of repetitions of the same commercials or (2) between 10-second and 30-second versions of the same commercials. Repetition and length variables are preferred because stimulus material remains constant and discrimination becomes more difficult to test. Our study was designed to make the discrimination test difficult.

The second objective of the study was *to show that recognition tests can adequately reveal that learning of a commercial has occurred*. Because recall is unusually stringent for accessing the memory of a commercial, one would predict that recognition scores would be higher than recall scores in general. The preceding two objectives basically address the quantitative aspect of memory testing.

The third objective was *to develop a qualitatively rich measurement*. The literature in psychology suggests that recognition has been measured mostly in an all-or-none fashion (i.e., is there recognition or not?). Our study involved multiple measures which examine recognition of *product class, brand name, claim,* and *overall package* of the brand.

On the basis of the three objectives, we tested the following hypotheses.

H_1: An increase in the repetition of a commercial will result in an increase in recognition and recall scores.

H_2: The recognition scores of 30-second commercials will be higher than those of their 10-second counterparts across all levels of repetition and distractors.

H_3: An increase in the number of distractors in a recognition test will lead to a corresponding decrease in recognition scores across all levels of repetition and message length.

H_4: Retention measured in terms of recall will be less than retention measured in terms of recognition across all levels of repetition and distractors.

METHOD

Experimental Design

The experimental design was a 2 (lengths of a commercial: 30-second and 10-second) × 2 (recognition tests: 5-alternative and 9-alternative tests) × 3 (levels of repetitions: 1, 2, and 4) split plot design. Lengths of commercials and levels of distractors were between-subject factors whereas level of repetition was a within-subject factor. Each subject saw 1, 2, or 4 repetitions of different commercials but saw only one length (i.e., either a 30-second or a 10-second commercial) and received either a 5-alternative (stimulus item with four distractors) or a 9-alternative (stimulus item with eight distractors) recognition test.

Subjects

Two hundred eleven undergraduate students volunteered to watch television for two half-hour periods and did so. Subjects signed up for convenient viewing periods; videotape treatments were assigned randomly to time slots, thereby randomly assigning subjects to cells. Cell sizes ranged from 48 to 59 over the 12 cells.

Procedure

The purpose of the study was disguised in tape-recorded instructions given to subjects before viewing began. Following is the text of instructions.

Welcome to our study. This study is being conducted to know what people like in a news show and why they select one news show over another. You are about to see portions of late night news shows taken from three network affiliates in three different cities. We have edited the news so that each station gets about ten minutes of time. In editing the news shows, we have tried to retain the flavor of the show, which means for each channel you will see the beginning and the end of the show, some news, weather, sports and commercials. Obviously, there will be some noticeable cuts and abruptness at certain spots, which we hope you will ignore. At the end of the show, you will fill out a questionnaire evaluating the news shows from the stations on such characteristics as personality of the news team, credibility of the news team, pace of the show and use of action and studio reporting.

In the second session, you will be shown three more news shows taken from cable affiliates. Thank you for participating in the study.

Viewing took place in groups of approximately 15 subjects. After viewing, subjects were asked to fill out an evaluation form comparing the news from the three stations. Subjects were asked to come back after two weeks to evaluate cable affiliate news shows. At this second session, all subjects were given an unaided recall test of commercials seen on the videotape. This was followed by the recognition test. Approximately one-half of the subjects who had watched a given tape received 9-alternative verbal and visual recognition tests; the other half received 5-alternative verbal and visual recognition tests. Upon completion of the tests subjects were debriefed and the purpose of the experiment was revealed.

Stimulus Materials

Commercials. Three 30-second predominantly informational commercials representing three low-involvement product categories (trash bags, frozen pie, salad dressing) had earlier been selected for use in the study from a pool of 200. Informational commercials were used because they would give subjects something to write about for recall and recognition of "claims." None of the selected commercials had ever been aired in the test city and none of the brands represented in the commercials were marketed in the test city, although both brands and commercials were real. These 30-second commercials were edited to create their 10-second versions. Judgment as to the informational nature of the commercials was made by a panel of three expert judges and was confirmed by a panel of 19 lay judges.

Overall videotape. Six 30-minute tapes were prepared with commercials embedded in the program material in six positions. These positions were between the news and weather and between the weather and sports of each of the three news shows. The first two and last two commercials on each tape were nonexperimental fillers used to avoid primacy and recency effects among the experimental commercials. No commercial was shown twice in a row. Each tape contained one experimental commercial at one repetition level, one at two, and one at four repetition levels. Fillers were all 30-second commercials and their position remained fixed in all six tapes. On the first three tapes all experimental commercials were 30 seconds long, whereas on the last three tapes all experimental commercials were 10 seconds long. Experimental commercials were rotated within the first three tapes and the last three tapes so that across these tapes the repetition levels and positions of the commercials were balanced. Thus a subject exposed to any of the first three tapes was exposed to one level of commercial length (:30) and all levels of repetition (1, 2, 4). Similarly, a subject exposed to any of the last three tapes received a different level of commercial length (:10) and all levels of repetition.

Measurements

Each subject received an unaided recall test first, followed by either 5- or 9-alternative verbal and visual recognition tests.

Recall test. Subjects were asked to recall the product category, brand name, and claim(s) for as many as possible of the commercials shown to them two weeks earlier.

Verbal recognition test. Subjects next completed sequential recognition tests of product category, brand name, and claim. Each subject was assigned to either a 5- or 9-alternative set of tests. First, subjects did a recognition task for a product category; on the next page, they were told the product category and asked to recognize the brand name from the given alternatives. On the next page they were given a number of claims and asked to select the correct claim. They were not told the correct brand name prior to the claim test.

Visual recognition test. Subjects were shown slides containing pictures of a number of packages (9- or 5-alternative recognition tests) in each of the three product categories and asked to identify the one they had seen in the news program. This was done because Rossiter (1976) showed that graphic assessment techniques revealed information stored in memory which was not shown by the verbal measures.

RESULTS

Table 1 reports the recognition scores and Table 2 the recall scores obtained in each experimental cell. All scores represent a composite of the three experimental commercials to avoid commercial-specific results and to increase generalizability of results. The entries in these tables represent the proportion of subjects who were able to recognize or recall the appropriate stimulus properties. Analysis of variance results for the recognition and recall data are summarized in Tables 3 and 4, respectively.[1] There are two error terms in each of these tables; the first error term relates to between-factor effects whereas the second error term relates to within-factor effects and the interactions with the between-factor effects.

---

[1]Several studies have shown that analysis of variance can be performed safely on nominal data if the number of observations per cell is large (≥ 30) (Benepe 1949; Child 1946; Cochran 1947; Mandeville 1969; Snedecor 1946). The recognition data were also analyzed by using a log-linear model; results were almost identical to the ANOVA shown. Data on verbal recognition were also transformed using the confidence ratings supplied by the subjects, i.e., a subject who was absolutely confident of his decision but recognized false object received a score of -3 and a subject who was right in his recognition decision and was absolutely confident of his decision received a score of $+3$, etc. Thus, the entire set of verbal recognition data was transformed to interval scale. (This practice is borrowed from psychology; see Pollack and Decker 1964 and Clarke 1964.) ANOVA on these data yielded the same results with minor differences. These corroborating tests seem to indicate that performing ANOVA in this case is appropriate.

Table 1
AGGREGATE RECOGNITION DATA

Length of commercial	:10						:30					
Number of distractors	4			8			4			8		
Number of repetitions	1	2	4	1	2	4	1	2	4	1	2	4
n	59	59	59	51	51	51	48	48	48	53	53	53
Product category proportion	.576	.763	.881	.392	.647	.961	.625	.875	.917	.604	.849	.962
σ_p	.06	.06	.04	.07	.07	.03	.07	.05	.04	.07	.05	.03
Brand name proportion	.475	.729	.814	.373	.549	.765	.479	.708	.854	.472	.547	.830
σ_p	.06	.06	.05	.07	.07	.06	.07	.07	.05	.07	.07	.05
Major claim proportion	.424	.627	.525	.333	.451	.471	.375	.583	.833	.509	.660	.793
σ_p	.06	.06	.07	.07	.07	.07	.07	.07	.05	.07	.07	.06
Visual identification proportion	.593	.881	.898	.373	.784	.804	.708	.813	.833	.547	.755	.925
σ_p	.06	.04	.04	.07	.06	.06	.07	.06	.05	.07	.06	.04

Results Pertaining to Discriminability of the Recognition Measure

The first objective of the study was to develop a recognition measure which would discriminate between various stimuli and would, in addition, not produce extremely high scores in all cases. An inspection of Table 1 shows that the lowest value is 33% for recognition of claims made in 10-second commercials to which subjects were exposed only once when their memory was tested on an 8-distractor (or 9-alternative) test, whereas the highest score is 96% for recognition of product category represented by 30-second commercials at four repetition levels when subjects were given an 8-distractor

Table 2
AGGREGATE RECALL DATA

Length of commercial	:10			:30		
Number of repetitions	1	2	4	1	2	4
n	110	110	110	101	101	101
Product category proportion	.155	.209	.736	.158	.446	.842
σ_p	.03	.04	-.04	.04	.05	.04
Brand name proportion	.018	.036	.182	.000	.050	.208
σ_p	.01	.02	.04	.00	.02	.04
Major claim proportion	.036	.091	.345	.069	.228	.485
σ_p	.02	.03	.05	.03	.04	.05

(or 9-alternative) test. Thus, the first objective is achieved in that scores do not cluster near an asymptote but rather are dispersed across a wide range of values.

Table 5 summarizes the data of Tables 1 and 2 so that the issue of discriminability can be examined further. The recognition test provides a broader range and a significantly greater variance for the brand name than does the recall test ($P < .01$; Hartley's F max test); the recall test gives a broader range and greater variance for product class and claim measures but the differences are not significant.

Practitioners have been concerned that recognition tests do not discriminate because the test encounters a ceiling effect due to its lack of rigor in challenging memory. If one is to test this issue, one should also consider the potential for a floor effect when testing recall due to the stringency of this memory test. To test for ceiling and floor effects, the bottom row of Table 5 shows the proportion of cells with scores above 80% or below 20%. Product, brand, and claim all show a greater proportion of extreme scores for recall (41.7%, 91.7%, 50%, respectively) than for recognition (50%, 25%, 8.3%, respectively). If there is a detriment to be associated with a clustering of extreme scores, it is more likely to occur in a recall test than in a recognition test.

To systematically demonstrate that the recognition tests developed herein are discriminating, recognition scores must be shown to differ significantly across number of repetitions and across length of commercials. These differences are discussed in the following section. Main effects are discussed first, followed by interaction effects.

Effect of repetition of commercials on recognition

Table 3
ANALYSIS OF VARIANCE TABLE FOR RECOGNITION SCORES

Source of variation	d.f.	Product recognition			Brand recognition			Claim recognition			Visual recognition		
		MS	F	P<	MS	F	P<	MS	F	P<	MS	F	P<
Distractors (D)	1	0.215	1.23	.269	1.197	6.26	0.013	0.099	0.63	0.428	1.365	7.74	0.006
Length (L)	1	1.634	9.34	.003	0.154	.80	0.371	3.722	23.60	0.001	0.227	1.29	0.257
D × L	1	0.209	1.20	.276	0.083	.44	0.510	1.058	6.71	0.010	0.308	1.75	0.188
Error	207	0.175			0.191			0.158			0.176		
Repetitions (R)	2	7.746	51.46	.001	7.033	31.93	0.001	3.309	12.41	0.001	5.640	33.87	0.001
R × D	2	0.403	2.68	.070	.277	1.26	0.285	0.087	0.33	0.720	0.472	2.83	0.060
R × L	2	0.284	1.89	.153	.071	.32	0.725	1.026	3.85	0.022	0.547	3.29	0.038
R × D × L	2	0.130	0.86	.423	.023	.11	0.900	0.224	0.84	0.432	0.098	0.59	0.557
Error	414	0.151			.220			0.267			0.167		

scores.[2] There is a significant main effect due to repetition (see Table 3) on product category recognition ($F = 51.64$; $P < .001$), brand name recognition ($F = 31.93$; $P < .001$), claim recognition ($F = 12.41$; $P < .001$), and visual recognition ($F = 33.87$; $P < .001$).

Repetition (see Table 4) has a significant effect on product ($F = 145.34$; $P < .001$), brand ($F = 31.51$; $P < .001$), and claim recall ($F = 79.98$; $P < .001$). These results are consistent with those of numerous advertising and verbal learning studies, and provide full support for H_1.

Effect of length of commercials on recognition scores. Length of commercial has a significant effect on recognition of product category ($F = 9.34$; $P < .003$) and claims ($F = 23.6$; $P < .001$), but does not have any significant effect on brand name recognition. Length of commercial also has no significant effect on visual recognition beyond one repetition.

A significant interaction is found between length and repetition for claim recognition ($F = 3.85$; $P < .02$). Figure 1 shows that in recognition of claims, 30-second commercials do better than 10-second commercials in all cases. This effect is only significant ($P < .05$) at the 4-repetition level, however. This finding suggests that an advertiser who is interested in getting claims across to

[2]Whenever there is a significant interaction between two factors, the main effects are qualified by testing the difference between the cell means of one factor at each level of the other factor using Tukey's ratio (1953).

the audience should use 30-second spots.

A significant interaction is found between repetition and length for visual recognition ($F = 3.29$; $P < .038$). Figure 1 also shows that 30-second commercials produced significantly higher visual recognition scores ($P < .05$) than 10-second commercials at the 1-repetition level but at two and four repetitions the differences are not significant. This finding suggests that for the purposes of achieving visual recognition, 10-second spots could be as effective as 30-second spots when a high-repetition media schedule is employed. The results thus offer only partial support for H_2.

Effect of number of distractors on recognition scores. Number of distractors has a significant effect on brand name ($F = 6.26$; $P < .013$) and visual recognition ($F = 7.74$; $P < .006$). However, it has no significant effect on product category recognition beyond one repetition.

There is a significant interaction between *repetition and distractor for product category recognition* ($F = 2.68$; $P < .07$). From Figure 2, we see that eight distractors did help reduce the recognition scores for product category at the 1- and 2-repetition levels but not at four. At the 1-repetition level, scores for the 4-distractor test were significantly higher than for the 8-distractor test ($P < .07$). However, there are no significant differences at two and four repetitions.

A significant interaction is found between *repetition and number of distractors for visual recognition* ($F = 2.83$; $P < .06$). Figure 2 also shows that eight distractors help reduce visual recognition scores more than four dis-

Table 4
ANALYSIS OF VARIANCE TABLE FOR RECALL SCORES

Source of variation	d.f.	Production recall			Brand recall			Claim recall		
		MS	F	P<	MS	F	P<	MS	F	P<
Length (L)	1	2.096	10.70	.001	.017	0.20	.652	8.30	12.83	.001
Error	209	0.196			.082			0.647		
Repetition (R)	2	22.549	145.34	.001	2.181	31.51	.001	43.794	79.98	.001
R × L	2	0.716	4.62	.010	0.039	0.56	.570	1.131	2.06	.128
Error	418	0.155			0.069			0.548		

Table 5

AGGREGATE TABLE COMPARING CERTAIN STATISTICS FOR RECALL AND RECOGNITION SCORES

	Recognition				Recall		
	Product	Brand	Claim	Visual	Product	Brand	Claim
Range	0.57	0.48	0.50	0.55	0.74	0.26	0.64
Mean	0.75	0.63	0.55	0.74	0.43	0.08	0.31
Standard deviation	0.18	0.17	0.16	0.16	0.29	0.09	0.22
% of cells (of 12) having a recognition or recall score:							
Above 0.80	50.0	25.0	8.3	50.0	16.7	0.0	0.0
Below 0.20	0.0	0.0	0.0	0.0	41.7	91.7	50.0

Figure 1

INTERACTIONS BETWEEN REPETITION AND LENGTH
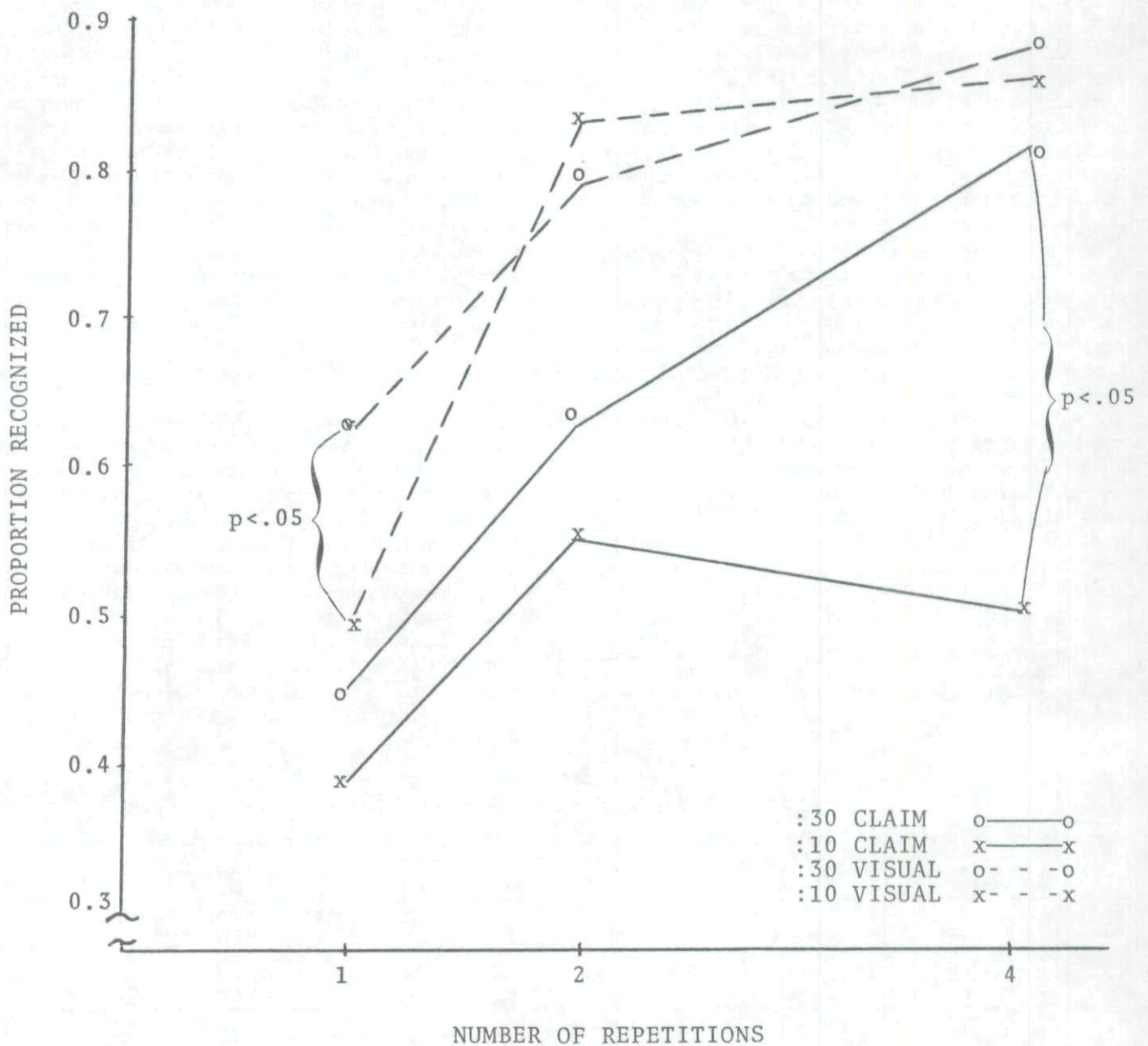
Figure 2
INTERACTIONS BETWEEN REPETITION AND DISTRACTOR

NUMBER OF REPETITIONS
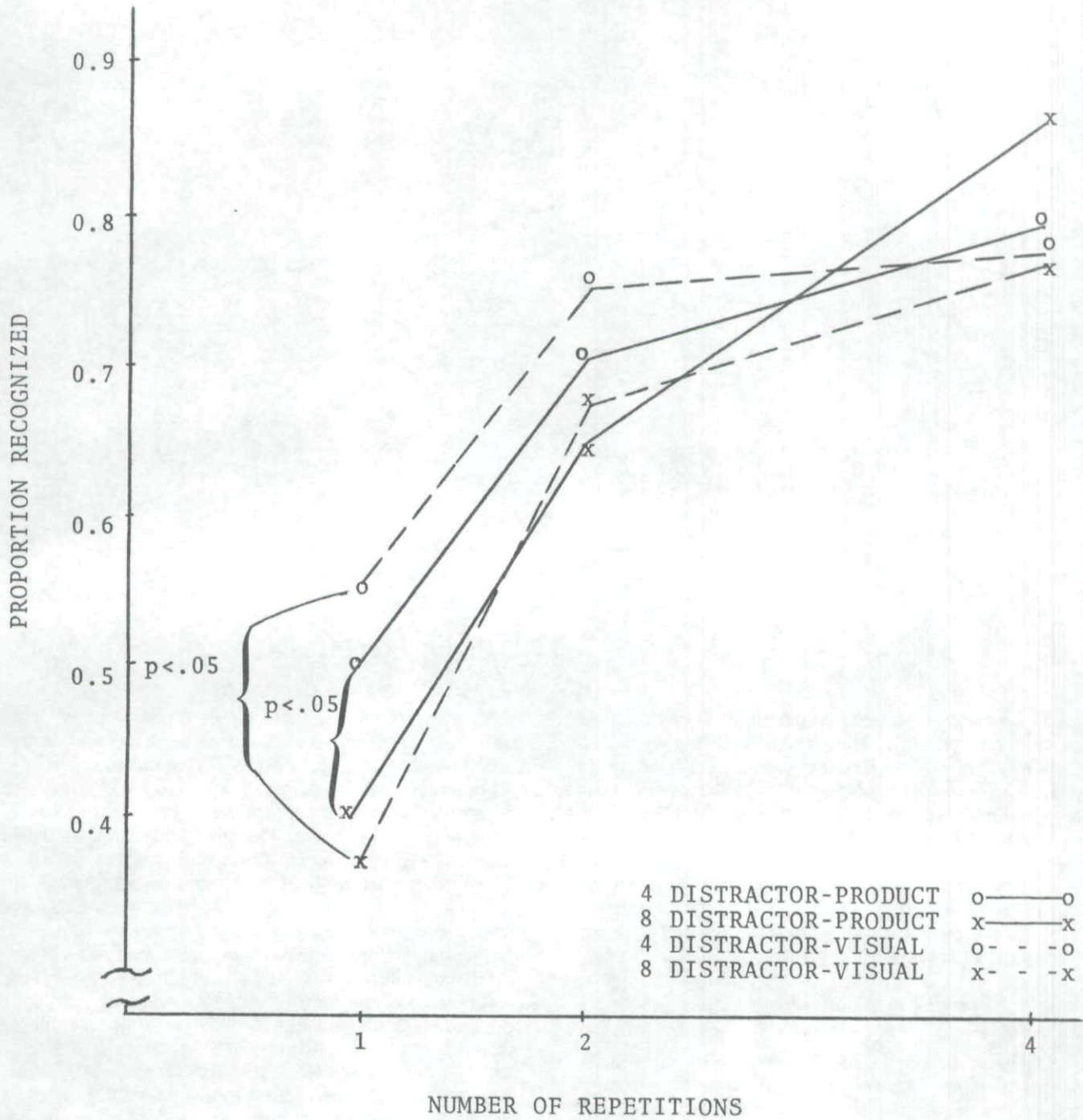
tractors at all repetition levels. However, Tukey's ratio for testing the difference between cell means reveals that the differences in visual recognition for 8- versus 4-distractor tests are significant only at the 1-repetition level ($P < .05$). Figure 2 suggests that as repetitions increase, the memory of which product class was advertised as well as the visual image of the advertised brand becomes

stronger and hence the effect of distractors disappears.

There is a significant interaction between *length and distractors for recognition of claims* ($F = 6.71$; $P < .01$). Number of distractors has a significant effect on claim recognition for 10-second but not for 30-second commercials. Figure 3 shows that when number of distractors increases from four to eight, the claim recog-

Figure 3
INTERACTION BETWEEN DISTRACTOR AND LENGTH
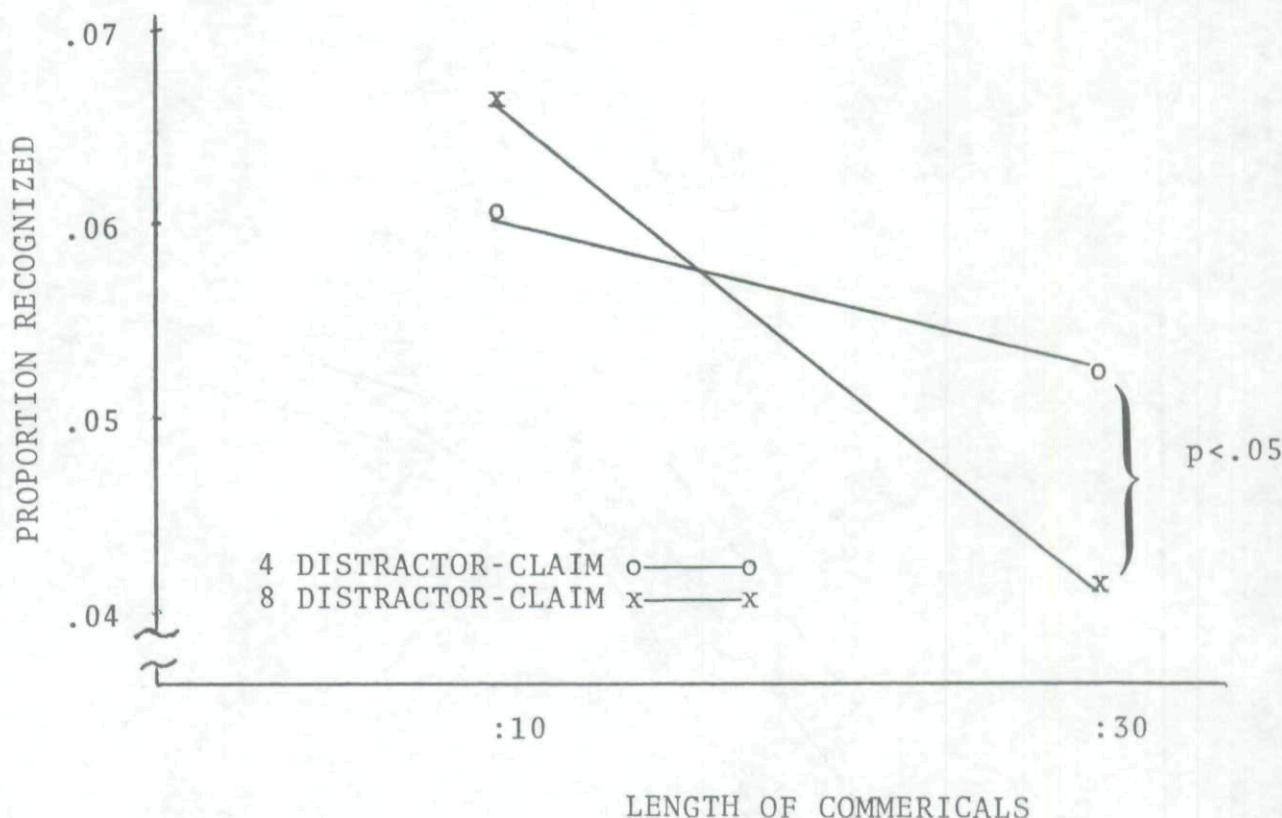
nition score for 10-second commercials goes down significantly ($P < .05$). However, the difference is not significant for 30-second commercials. This finding implies that distractors are more effective in reducing claim recognition scores for 10-second spots than for 30-second spots. Claims may be remembered less well in a cluttered purchase environment.

In sum, an increase in the number of distractors in a recognition test leads to a corresponding decrease in recognition scores, but it does not do so uniformly. Hence, the results offer only a *partial support to H_3*.

Results Pertaining to the Sensitivity of Recognition Measures

Sensitivity of a measure here refers to its capability to access memory for a commercial. How "sensitive" is recognition in comparison with recall? Are recall measures too stringent a test for memory of a commercial? The answers to these questions require a comparison of recognition and recall scores. Tables 1 and 2 afford such a comparison.

These tables show that when we are examining similar levels of length of commercial (30-second or 10-second), levels of repetition (1, 2, and 4), and distractor levels (8 or 4), recall scores are significantly lower than

recognition scores in all but two of 36 comparisons.[3] This difference is greatest for brand name recall and recognition. As an example, consider 30-second commercials. At one repetition level, recall scores are 0.0% for both 8- and 4-distractor tests, whereas recognition scores in corresponding cells are 47% and 48%, respectively. Similarly, for four repetitions and distractors, recall is 26% in comparison with a recognition score of 83%; comparable scores are 15% and 85% for the 4-distractor test. These results provide *ample support for H_4*.

Not only are recognition scores higher than recall scores in all equivalent cases, but recognition scores for 10-second commercials are also higher than recall scores for 30-second commercials in 15 of 18 cases when numbers of repetitions and distractors are the same ($p < .05$; Z test). These data show that recall measures mask the existence of some memory trace for a commercial, whereas recognition tests are more sensitive to revealing the oc-

---

[3]The McNemar test for related samples was used to test the difference between recall and recognition scores. All differences were found significant at a probability level of 0.05 or below, except for product category recognition of 30-second commercials at 4-repetition and 4-distractor levels and claim recognition of 10-second commercials at 4-repetition and 8-distractor levels, where differences between recall and recognition scores were not significant.

currence of learning and memory.

In the same vein, recognition scores are higher than recall scores with fewer repetitions in 14 of 24 cases ($p < .05$; McNemar's test of related samples). Recall scores only exceed the recognition scores when four versus one repetitions are compared.

A final comparison of recall and recognition considers whether the two tests make similarly ordered predictions. When the data are ranked in order of cell scores, Spearman's rho shows a significant correlation between recall and recognition for product (rho = .923; $P < .01$), brand (rho = .916; $P < .01$), and claim (rho = .748; $P < .01$). Recall and recognition lead to a similar ordering of scores across test cells; that is, they make similar predictions.

These strong correlations are consistent with the dual process hypothesis (Anderson and Bower 1974; Kintsch 1974) which implies that both recall and recognition involve similar processes.

Rank order correlations between product recognition, brand recognition, claim recognition, and visual recognition tests were also calculated. The correlations range from rho = 0.72 (between claim recognition and visual recognition) to rho = 0.93 (between brand recognition and visual recognition). All correlation coefficients are significant at $P < .01$, suggesting that the four recognition tests are measuring the same underlying construct.

DISCUSSION

Our study had three major objectives:

1. To develop a recognition measure that would not yield consistently high scores and would discriminate across stimuli.
2. To show that learning of commercial information occurs even though recall measures may be unable to tap this learning.
3. To develop a qualitatively rich measurement.

To achieve the first objective, a large number of distractor items and a two-week delay were used to make the recognition task difficult, and a forced-choice recognition test was used so that response bias (noting set tendency) would not exert much influence on the observed choices. Increasing the number of distractors in the recognition test did help reduce the recognition scores for brand name product category and visual recognition. However, an increase in distractors did not prove effective in lowering recognition scores for product category recognition and visual recognition at higher repetition levels. Similarly, an increase in distractors failed to reduce claim recognition for the 30-second version of the commercials.

It seems that as the number of repetitions increases, the memory for product category and visual recognition becomes stronger and the recognition task must be very difficult to lower the recognition scores. An increase in distractors from four to eight was not able to make the recognition task difficult enough across all conditions.

The data show a wide range of learning across the different levels of repetition and message lengths. In addition, recognition scores are less clustered around some high value than are recall scores around some low value. Thus a recognition test can be developed which separates good from poor learning commercials. In our study, a delayed recognition test and a large number of distractors used in a forced-choice test led to such a separation. The first objective is therefore met.

A more general issue related to the first objective is the construction of a discriminating test. Though we used two means to this end (delayed testing and increased numbers of distractors), other means are also available. The books by Bloom (1956) and Gage (1963) are useful references for construction of educational tests; those of Nunnally (1978) and Kerlinger (1964) are useful in the development and testing of discriminating items. An important area to pursue in future recognition tests is the similarity of choice items. The works cited include various ways of measuring similarity of test items.

The data also indicate that learning can be measured in situations where recall does not show that learning has occurred. In many of the cells there is a considerable amount of recognition even though recall scores are zero or close to zero. This finding indicates that learning of commercial information occurs even though recall measures are unable to tap this learning; thus the second objective of the study is achieved. One should keep in mind that the data represent a two-week delayed test. Shorter delay periods might result in higher scores although the relations presented above should continue to exist. We are examining the issue of varying lengths of decay periods in a second study.

Finally, though most studies in the psychology and advertising literatures have measured recognition in an all-or-none fashion, we measured it on four dimensions: product, brand, claim, and visual package. The study thus provides a qualitatively richer measure of recognition. Though this task was not particularly profound, it had not been previously done and it shows advertisers that rich data can be obtained in a recognition test.

This richness can be enhanced more formally by adoption of a multitrait, multimethod philosophy (Campbell and Fiske 1959; Heeler and Ray 1972). One can develop a multiple-item recognition test which evaluates various stimulus components. This logic led to the multipart test used herein which examined product, brand, claim, and package, but one could expand it by testing several claims or product attributes and could then test their convergent and discriminant validity. Though the third objective of the study is met, a more formal process could aid future researchers to achieve even richer measures.

Having met our three objectives, we can consider the implications of the data for media strategy, awareness testing, and copy testing.

Media Strategy

If a recognition level of learning is sufficient (e.g., a case in which the product class elicits low involvement

in general and/or brand choice decisions are typically made in store), a reduced media schedule also may be sufficient. Recognition scores are generally higher than recall scores even when the recognition test is conducted under a situation of less repetition. If the advertiser's goal is (for example) 70% awareness of a particular claim, this level may have been reached much earlier in the media plan than had previously been thought. Sparser media plans may be sufficient for consumer convenience products because a recognition level of learning may be sufficient to aid the decision-making process. Indeed, the entire concept of awareness may need to be reconsidered in light of the differences between recall and recognition.

Media planning is an area where managers admit to huge gaps of knowledge in terms of causal relationships. The question of how much repetition is enough is answered with a general lack of specificity. As a result, media planners may have a tendency to overspend to ensure success for clients. If a recognition level of learning is sufficient, these planners may be overspending by large amounts.

The recall/recognition issue also affects media strategy in the area of message length, and popular notions about length should be reconsidered. If the goal of the advertiser is brand name learning, then 10 seconds seem to be sufficient. In our study, brand name recognition of the 10-second commercials outperformed brand name recall of the 30-second across all cases of comparable repetitions or comparable distractors. Brand name recognition of the 10-second commercials also exceeded brand name recall of the 30-second regardless of the number of repetitions of the latter.

Brand name recognition of the 10-second commercials also was not significantly different from recognition of the 30-second. This finding supports the value of 10 seconds in achieving cost-efficient brand name learning. The difference in recall of brand name between 10 and 30 seconds is also not significant, but with the recall test the data could be interpreted as being biased by a floor effect. Such a bias does not exist in the recognition comparison.

Many commercials have brand name learning as their primary goal. In these cases 10 seconds may be sufficient. Coincidentally, commercials for which brand name learning (versus claim learning) is the goal are generally for commodity-type products for which people usually have a lower involvement, choice decisions are made in store, and a recognition level of learning may be sufficient.

In contrast, if the goal of the advertiser is claim learning, 30-second are better than 10-second commercials, especially at higher repetition levels. Here claim learning (in terms of recognition) is better promoted by 30-second commercials. The strategic implications for claim learning, though, are not as straightforward as those for brand name.

Current practice is to use 10 seconds to maintain

learning that has been developed through the use of 30 seconds. The data suggest that in some cases 10-second commercials could be used earlier in the campaign.

In our study we examined recall and recognition of television commercials after a two-week delay. Memory has been shown to exist after this delay even though the recall test could not measure it. Advertisers should reconsider how much learning can take place under conditions of lower repetition and/or shorter messages, and whether a recall level of learning is really necessary for their products.

Awareness Testing

One can also consider our study in terms of the development of a recognition-based awareness test. Such tests currently are available for print (Starch) and television (Bruzzone Research Company). Though the purpose of our study was purely to examine the concept of recognition, one can also derive implications from the study for applied commercial testing. The method could be implemented via a mall intercept process, and recognition scores could be standardized on the basis of the length of the campaign, its heaviness (GRP levels), and the amount of time elapsed since its curtailment. Various values of these three variables could be tested through matched experimental markets.

In a separate study we are examining the issue of elapsed time between stimulus receipt and measurement. When those data are included with the current set, we perhaps can make more definitive statements in terms of developing an applied recognition test. The applicability of the current findings is limited because of the single value of the delay variable (two weeks) and the laboratory setting.

Copy Testing

A fundamental task of copy testing is to discriminate between commercials. We did not attempt to discriminate between commercials; rather, we believed a more difficult task would be to discriminate between lengths and repetition levels of the same commercial. This task was collapsed across three product classes to give more generalizable results. Ultimately, our work should be replicated across a variety of commercials.

A second task is to provide insight to copy writers and artists. Although neither recall nor recognition tests are designed to give in-depth information in this area, both can provide a general diagnostic check as to whether brand name and major claim are communicated. Neither test does this well, however, and both should be used in conjunction with other more appropriate tests if diagnostic information is desired.

Given that the two tests are equally good (or poor) at copy-testing tasks, one should choose between them on the basis of the type or strength of memory in one's learning model. Our study, having shown the utility of recognition tests, has also made this choice more feasible for practitioners.

Our study was a laboratory experiment designed to optimize internal validity issues. The next step would be to do a field study of recognition to achieve requisite external validity. Though we gained some level of external validity by imbedding commercials in program content and disguising the purpose of the experiment, the goal was internal validity; consequently marketplace variables such as competitive spending levels were not considered, and, indeed, no competitive commercials were shown. If field tests were to support our laboratory results, the implications to the advertising industry could be far reaching and could include reduced budgets, shorter messages, and increased clutter.

REFERENCES

Anderson, J. R. and G. H. Bower (1972), "Recognition and Retrieval Processes in Free Recall," *Psychological Review*, 79, 97–123.

———— and ———— (1974), "A Propositional Theory of Recognition Memory," *Memory and Cognition*, 2 (3), 406–12.

Appel, V. and M. L. Blum (1961), "Ad Recognition and Respondent Set," *Journal of Advertising Research*, 1, 13–21.

Atkinson, R. C. and J. F. Juola (1973), "Factors Influencing Speed and Accuracy of Word Recognition," in *Attention and Performance*, Vol. 4, S. Kornblum, ed. New York: Academic Press.

Beals, H., M. B. Mazis, S. C. Salop, and S. Staelin (1981), "Consumer Search and Public Policy," *Journal of Consumer Research*, 8 (1), 11–22.

Benepe, O. J. (1949), *The Sensitivity of t and F to Departures from Normality*, Master's thesis, University of Washington.

Bettman, J. R. (1979), "Memory Factors in Consumer Choice: A Review," *Journal of Marketing*, 43 (2), 37–53.

Bloom, B. (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I, Cognitive Domain*. New York: David McKay.

Campbell, D. T. and D. W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, 81–105.

Child, I. L. (1946), "Children's Preference for Goals Easy or Difficult to Obtain," *Psychological Monograph*, No. 280.

Clancy, K. J., L. E. Ostland, and G. A. Wyner (1979), "False Reporting of Magazine Readership," *Journal of Advertising Research*, 19 (5), 23–30.

Clarke, F. R. (1964), "Confidence Ratings, Second Choice Responses, and Confusion Matrices in Intelligibility Tests," in *Signal Detection and Recognition by Human Observers*, J. A. Swets, ed. New York: John Wiley & Sons, Inc., 620–48.

Cochran, W. G. (1947), "Some Consequences When the Assumptions for the Analysis of Variance Are Not Met," *Biometrics*, 3, 22–38.

Craik, F. I. M. (1971), "Age Differences in Recognition Memory," *Quarterly Journal of Experimental Psychology*, 23, 316–23.

Dallett, K., S. G. Wilcox, and L. D'Andrea (1968), "Picture Memory Experiments," *Journal of Experimental Psychology*, 76 (2), 312–20.

Davis, R., N. S. Sutherland, and B. R. Judd (1961), "Information Content in Recognition and Recall," *Journal of Experimental Psychology*, 61, 422–29.

Eagle, M. and E. Leiter (1964), "Recall and Recognition in

Intentional and Incidental Learning," *Journal of Experimental Psychology*, 68 (1), 58–63.

Gage, N. (1963), *Handbook of Research on Teaching*. Skokie, IL: Rand McNally.

Goldstein, A. G. and J. E. Chance (1970), "Visual Recognition Memory for Complex Configurations," *Perceptions and Psychophysics*, 9, 237–41.

Haber, R. N. (1970), "How We Remember What We See," *Scientific American*, 222 (May), 104–12.

Hansen, F. (1981), "Hemispheral Lateralization: Implications for Understanding Consumer Behavior," *Journal of Consumer Research*, 8 (1), 23–36.

Haskins, J. B. (1964), "Factual Recall as a Measure of Advertising Effectiveness," *Journal of Advertising Research*, 4 (March), 2–8.

Heeler, R. M. and M. L. Ray (1972), "Measure Validation in Marketing," *Journal of Marketing Research*, 9 (November), 361–70.

Kerlinger, F. N. (1964), *Foundations of Behavioral Research*. New York: Holt, Rinehart and Winston, Inc.

Kintsch, W. (1970), "Models for Free Recall and Recognition," in *Models of Human Memory*, D. A. Norman, ed. New York: Academic Press.

———— (1974), *The Representation of Meaning in Memory*. Potomoc, MD: Lawrence Erlbaum Associates.

Klatzky, R. L. (1980), *Human Memory: Structures and Processes*. San Francisco: W. H. Freeman and Company.

Krugman, H. E. (1971), "Brain Wave Measures of Media Involvement," *Journal of Advertising Research*, 11 (February), 3–10.

———— (1972), "Why Three Exposures May Be Enough," *Journal of Advertising Research*, 12 (December), No. 6, 11–14.

———— (1977a), "Memory Without Recall, Exposure Without Perception," *Journal of Advertising Research*, 17 (4), 7–12.

———— (1977b), *Marketing News*, March 25, Vol. X, No. 18.

———— (1979), "Low Involvement Theory in the Light of New Brain Research," in *Attitude Research Plays for High Stakes*, J. C. Maloney and B. S. Silverman, eds. Chicago: American Marketing Association.

Lucas, D. B. (1960), "ABCs of ARF's PARM," *Journal of Marketing*, 25, 9–20.

———— and S. H. Britt (1963), *Measuring Advertising Effectiveness*. New York: McGraw-Hill Book Company.

Mandeville, G. K. (1969), *A Monte Carlo Investigation of the Adequacy of Standard Analysis of Variance Test Procedures for Dependent Binary Variates*, Ph.D. thesis, University of Minnesota.

Mandler, G. (1972), "Organization and Recognition," in *Organization of Memory*, E. Tulving and W. Donaldson, eds. New York: Academic Press.

———— (1980), "Recognizing: The Judgment of Previous Occurrence," *Psychological Review*, 87, 252–71.

————, Z. Pearlstone, and H. S. Koopmans (1969), "Effects of Organization and Semantic Similarity on Recall and Recognition," *Journal of Verbal Learning and Verbal Behavior*, 8, 410–23.

Martin, E. (1975), "Generation-Recognition Theory and the Encoding Specificity Principle," *Psychological Review*, 82, 150–3.

Nunnally, J. C. (1978), *Psychometric Theory*. New York: McGraw-Hill Book Company.

Pollack, I. and L. R. Decker (1964), "Confidence Ratings, Message Reception and Receiver Operating Characteristic,"

in *Signal Detection and Recognition by Human Observers,* J. A. Swets, ed. New York: John Wiley & Sons, Inc., 592–608.

Reder, L. M., J. R. Anderson, and R. A. Bjork (1974), "A Semantic Interpretation of Encoding Specificity," *Journal of Experimental Psychology,* 102, 648–56.

Rossiter, J. R. (1976), "Visual and Verbal Memory in Children's Product Information Utilization," in *Advances in Consumer Research,* Vol. III, B. B. Anderson, ed. Cincinnati: Association for Consumer Research, 523–27.

Santa, J. L. and L. L. Lamwers (1974), "Encoding Specificity: Fact or Artifact?" *Journal of Verbal Learning and Verbal Behavior,* 13, 412–23.

—— and ——(1976), "Where Does the Confusion Lie? Comments on the Wiseman and Tulving Paper," *Journal of Verbal Learning and Verbal Behavior,* 15, 53–7.

Shepard, R. N. (1967), "Recognition Memory for Words, Sentences and Pictures," *Journal of Verbal Learning and Verbal Behavior,* 6, 156–63.

—— and J. J. Chang (1963), "Forced Choice Tests of Recognition Memory Under Steady State Conditions," *Journal of Verbal Learning and Verbal Behavior,* 2, 93–101.

Snedecor, G. W. (1946), *Statistical Models,* 4th ed. Ames, IA: State College Press.

Standing, L., J. Conezio, and R. M. Haber (1970), "Perception and Memory for Pictures: Single Trial Learning of 2500 Visual Stimuli," *Psychological Science,* 19 (2).

Strong, E. K. (1912), "The Effect of Length of Series Upon Recognition Memory," *Psychological Review,* 19, 447–62.

Tukey, J. W. (1953), "The Problem of Multiple Comparisons," Princeton University (Ditto), 396 p.

Tulving, E. and D. M. Thomson (1973), "Encoding Specificity and Retrieval Processes in Episodic Memory," *Psychological Review,* 80, 352–73.

Tversky, B. (1973), "Encoding Processes in Recognition and Recall," *Cognitive Psychology,* 5, 275–87.

Underwood, B. J. (1965), "False Recognition Produced by Impact Verbal Responses," *Journal of Experimental Psychology,* 70 (1), 122–9.

—— and J. S. Freund (1968), "Errors in Recognition, Learning and Retention," *Journal of Experimental Psychology,* 78 (1), 55–63.

Wiseman, S. and E. Tulving (1976), "Encoding Specificity: Relation Between Recall Superiority and Recognition Failure," *Journal of Experimental Psychology: Human Learning and Memory,* 2, 349–61.

Zielske, H. (1959), "The Remembering and Forgetting of Advertising," *Journal of Marketing,* 23 (January), 239–43.