A distractor-free test of recognition and false recognition

RONALD LEY and KATHLEEN LONG
State University of New York at Albany, Albany, New York

Following the presentation of 36 words, one at a time, a distractor-free test of word recognition (a single-item test trial in which the 36 targets were presented prior to the 36 distractors) or a traditional target-distractor discrimination test (a single-item test trial in which the 36 targets and 36 distractors were randomly ordered) was given to 137 college students. Contrary to the long-standing assumption that distractors are necessary in tests of recognition in order "to keep the subject honest," the distractor-free test resulted in recognition performance (hits and serial-position frequency patterns) almost identical with that of the conventional discrimination test for high-, medium-, and low-frequency words. For the low-frequency words (words presumed to be unfamiliar to the subjects), false recognitions were fewer under the distractor-free test than under the discrimination test.

Wallace (1980) made an interesting methodological distinction between *recognition-discrimination* memory (the traditional measure of recognition in which target items are tested in the context of randomly dispersed distractors) and *recognition* memory (the infrequently used measure of recognition in which target items are tested in the absence of distractors; e.g., Tulving & Watkins, 1973; Underwood, 1972; Wallace, 1978; Wallace, Sawyer, & Robertson, 1978). This distinction is important for practical and theoretical reasons because it provides operational definitions that separate clearly the general common-language conception of recognition (knowing or feeling that a present object has been met before) from the limited conception of recognition contrived by experimental psychologists (discriminating between targets and distractors).

The importance of this distinction may have been obvious for a long time, but the assumption that tests of recognition must include distractors in order "to keep the subject honest" (e.g., Baddeley, 1976; Crowder, 1976; Klatzky, 1975; Woodworth, 1938) may have inhibited investigators from studying pure recognition memory, and encouraged instead the study of synthetic recognition-discrimination memory. The effect of this assumption in directing the course of research in memory is great, but its effect on misdirecting the course of research may be even greater. As Wallace (1980) concluded in the final paragraph of his essay;

A requirement for including distractors on recognition tests is simply not essential to the concept of recognition memory. At present, the estimates of an individual's capacity to recognize a prior experience are determined to a large extent by the experimenter's decisions regarding the number of distractors and the nature of distractors. For

the recognition memory implied by the description of *knowing or feeling that the present object has been experienced before*, test distractors are conceptually irrelevant and procedurally unnecessary. Even worse, their protracted use may be limiting the full understanding of recognition memory. (p. 702)

This conclusion is based in part on the results of experiments (Wallace, 1978; Wallace, Sawyer, & Robertson, 1978) in which the number of target words recognized following the conventional recognition-discrimination procedure (study trial followed by a target-distractor discrimination-test trial) did not differ from the number of words recognized when the test trial did not contain distractors. These results obtained for both high- and low-frequency words, and regardless of whether or not the subjects were instructed that all the test words were on the study list.

Although the effects of distractor variables are of interest in their own right, primarily in their potential for producing false recognitions, the distractor-free tests of recognition used by Wallace et al. (1978) precluded the measurement of false recognitions and thus limited their scope exclusively to recognition. The primary purpose of the present study was to test a new method for measuring both recognition memory (a count of correctly identified targets free of distractors) and false recognition (a count of distractors incorrectly identified as targets). The method is based on a procedure in which a study list of targets is followed by a test list in which all of the targets are presented and responded to prior to the presentation of the distractors. This simple segregated arrangement of targets and distractors on the test list requires only that each item of the list be presented one at a time and that the subject be required to respond "old" or "new" during a single presentation. If this single-item test procedure is followed, the subject shall have responded to all of the targets before having seen any distractors, thus preclud-

---

Requests for reprints should be sent to Ronald Ley, ED 228, State University of New York at Albany, 1400 Washington Ave., Albany, NY 12222.

ing any effects of discrimination that might result from distractor variables.

Although the single-item test used here differs from the embedded-item test used by Wallace (1978), recognition scores from these studies are suitable for comparison; Ley and Karker (1982) found that for the same list of targets and distractors, single-item tests and embedded-item tests yield equivalent recognition scores.

The secondary purpose of the present study was to determine whether the proposed innovation in the method for measuring recognition and false recognition applies to words with high-, medium-, and low-frequency ratings under conditions of group administration. If it is true that distractors are necessary in tests of recognition in order to keep the subjects honest, then the anonymity (the absence of a one-to-one exchange between the subject and the experimenter) that results from testing subjects in a relatively large intact group should favor "dishonesty" to a greater extent than would be expected when testing subjects individually. In addition to comparisons with Wallace's (1978) individually obtained recognition scores, the present study provides for the analysis of false recognition scores under two modes of testing (i.e., distractor-free tests and discrimination tests) within intact groups.

In this experiment, recognition scores from distractor-free tests were compared with scores from conventional discrimination tests. The comparisons involved three separate lists of targets: one composed of high-frequency words, one composed of medium-frequency words, and one composed of low-frequency words (words rarely identified as words). In addition, false-recognition scores from a target-free test were compared with those obtained from the conventional discrimination test. The target-free test consisted of the presentation of the distractors after all of the targets of the distractor-free test had been presented and responded to.

METHOD

Subjects

The subjects, 137 students at the State University of New York at Albany, were unpaid volunteers recruited from three classes: one ($N=54$) for testing the high-frequency words, one ($N=51$) for testing the medium-frequency words, and one ($N=32$) for testing the low-frequency words. The recruitment took place at the beginning of regularly scheduled class periods.

Materials and Procedure

The words for the three recognition tests were selected from Ley and Tesiny's (1975) list of words and paralogs assessed for associative frequency (a'). Each list consisted of 76 English language words (36 targets, 4 buffers, and 36 distractors). The high-frequency words were selected from the upper third of Ley and Tesiny's list, the medium-frequency words from the middle third of the list, and the low-frequency words from the bottom third. Since the low-frequency words of the present study were unknown to the experimenters, they are comparable to the very-low-frequency words used by Mandler, Goodman, and Wilkes-Gibbs (1982).

The study trials of the recognition procedure for the high-, medium-, and low-frequency words were the same. The 36 targets and 4 buffers of each list were displayed on a film screen via a Kodak Carousel slide projector, 1 word per slide at a 4-sec presentation rate controlled by

the internal timer of the projector. Two of the buffers were displayed before the list of 36 targets and 2 after the list.

During a 2-min period that intervened between the study and test phases, the subjects were instructed to perform simple arithmetic computations (the multiplication of three-digit numbers by three-digit numbers).

Prior to the study phase, one of two sealed test booklets was distributed randomly to each subject, after which the experimenter read aloud the detailed instructions that, in essence, told the subjects that the experiment consisted of three phases: (1) a study phase, in which they would see a list of words to be remembered, each word presented one at a time; (2) a computational phase, in which they would be required to do some multiplication exercises; and (3) a test phase, in which they would find in their sealed booklets a single word in the middle of each page with the words "old" and "new" in the bottom corners of each page, and that if the word in the middle of the page was one they remembered seeing during the study phase, they should encircle the word "old," and if it was not, they should encircle the word "new." The subjects were also instructed to go through the booklet once in a sequential order beginning with the first page, and not to refer to previously marked pages.

The study booklets consisted of two different formats, that is, two different arrangements of targets and distractors. Half the booklets (the distractor-free test) contained the 36 targets on the first 36 pages of the booklet and the 36 distractors on the last 36 pages. The other half contained a random mix of targets and distractors (the discrimination test), with the single restriction that the order of the targets and distractors be the same as that of the distractor-free test. There were no time restrictions during the testing phase, but all subjects completed the test booklets within 5 min.

RESULTS

Hits

The mean number of hits (correctly identified high-, medium-, and low-frequency target words) for the distractor-free and discrimination tests of recognition are given in Table 1. A cursory examination of these data makes it clear that the observed differences between the means of the distractor-free and discrimination tests were very small. For high-frequency words, the difference was .89 [$t(52) = .54, p > .05$]; for medium-frequency words, the difference was .22 [$t(49) = .17, p > .05$]; and for low-frequency words, the difference was .38 [$t(30) = .22, p > .05$].

In order to compare the data of the present study with those of Wallace et al.'s (1978) Experiment 2, the means from both studies were converted to proportion scores by dividing the means of the correct recognitions by the total number of targets on their respective lists. These mean proportions, which are listed in Table 2, indicate that the overall performance means of the present study ($M = .80$ for the distractor-free test and $M = .79$ for the discrimination test) are almost identical to those of Wallace et al.

Table 1
Mean Number of Words Correctly Identified (Hits)
Under Two Tests of Recognition

Word Frequency	Test of Recognition			
	Distractor-Free		Discrimination	
	Mean	SD	Mean	SD
High	28.59	6.42	27.70	5.33
Medium	29.67	5.30	29.45	3.81
Low	28.12	4.99	28.50	4.43

Table 2
Mean Proportions (Hits/Total Targets) of Words Recognized Under Two Tests of Recognition from the Present Study and from Wallace, Sawyer, and Robertson's (1978) Experiment 2

Word Frequency	Wallace et al., Experiment 2		Present Experiment	
	Distractor-Free	Target-Distractor Discrimination	Distractor-Free	Target-Distractor Discrimination
High	.75	.73	.79	.77
Medium			.82	.82
Low	.87	.87	.78	.79
Grand Mean	.81	.80	.80	.79

(M = .81 and M = .80, respectively). A direct comparison of the high- and low-frequency means do not, however, match as closely. The mean proportions of high-frequency targets of the present study (.79 and .77) were larger than those of Wallace et al. (.75 and .73), whereas the mean proportions of low-frequency targets of the present study (.78 and .79) were smaller than those of Wallace et al. (.87 and .87). It is important to note that the mean proportions of medium-frequency words (.82 for both tests) correspond more closely to Wallace et al.'s low-frequency proportions.

Although the use of intact groups precludes a determination of the statistical significance of the differences among the means of the high-, medium-, and low-frequency targets recognized, the word-frequency paradox (better recognition of low-frequency than high-frequency words) reported by Wallace et al. (1978) does not appear to obtain for all levels of word frequency of the present study. That is, the observed differences between the high-frequency and medium-frequency words (.79-.82) favor the medium-frequency words (.77-.82), but the observed differences between the high-frequency and low-frequency words (.79-.78 and .77-.79, respectively) are virtually nil. It is important to note, however, that the pattern of hits of the present study resembles the pattern of hits for high-, low-, and very-low-frequency words under the integration condition of Mandler et al.'s (1982) Experiment 2. As pointed out earlier, the low-frequency words of the present study and the very-low-frequency words used by Mandler et al. were essentially "nonwords."

Serial-Position Curves

Additional evidence of the comparability of the distractor-free and discrimination tests is given in Figures 1, 2, and 3, plots of the serial-position curves for proportions of high-, medium-, and low-frequency hits for each test. Interest here did not center on the serial-position effect; use of a single order of presentation confounds list position with word memorability. Examination of these figures indicates that each has a unique pattern, but that the pattern within each figure is essentially the same for both the distractor-free and the discrimination tests. Whatever variables account for the differences in the recognizability of the targets, both tests appear to be equally sensitive to them.

False Recognitions

The mean number of false recognitions of high-, medium-, and low-frequency words under the target-free and discrimination tests are given in Table 3. For high-frequency words, the mean number of false recognitions under the discrimination test (M = 4.41) was not significantly greater than the mean under the target-free test (M = 3.26) [$t(52)$ = 1.05, $p > .05$]; for the medium-
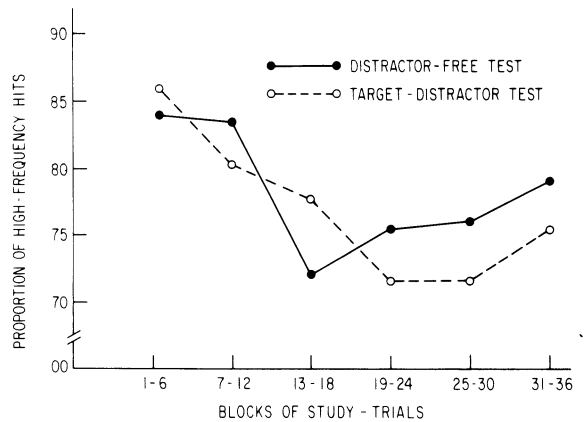

Figure 1. Serial-position curves of the proportions of high-frequency words recognized (hits) under the distractor-free and target-distractor discrimination tests.
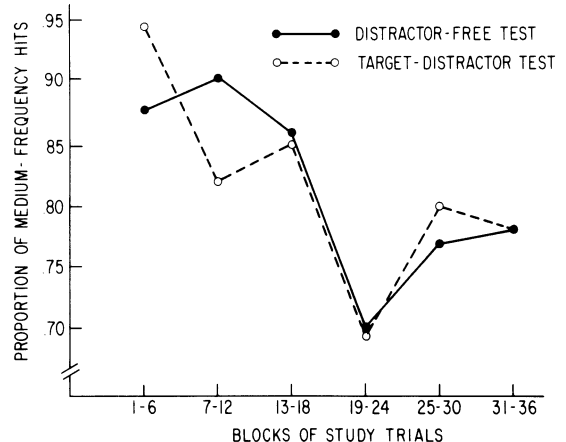

Figure 2. Serial-position curves of the proportions of medium-frequency words recognized (hits) under the distractor-free and target-distractor discrimination tests.
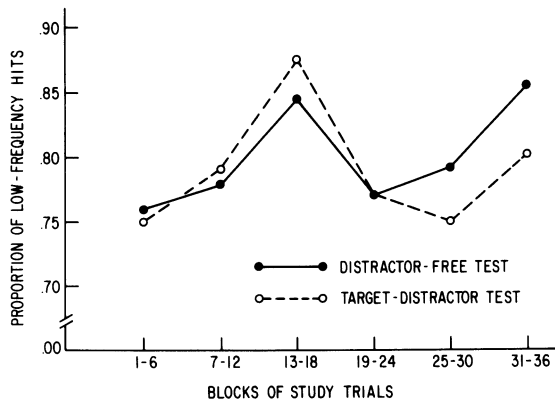
Figure 3. Serial-position curves of the proportions of low-frequency words recognized (hits) under the distractor-free and target-distractor discrimination tests.

frequency words, the mean of false recognitions under the discrimination test ($M = 8.17$) was not significantly different from the mean under the target-free test ($M = 5.63$) [$t(86) = 1.18, p > .05$]; but for the low-frequency words, the mean of false recognitions under the discrimination test ($M = 9.69$) was significantly greater than the mean under the target-free test ($M = 5.94$) [$t(30) = 2.27, p < .05$].

The use of intact groups precludes a test of the statistical significance of either the differences among false recognitions of high-, medium-, and low-frequency words, or the interaction between word frequency and the two tests of recognition.

DISCUSSION

The findings of the present study support Wallace's (1980) thesis that tests of recognition memory do not require distractors. The virtually identical mean recognition scores under the distractor-free and discrimination tests make it clear that distractors are not necessary in order to keep the subject honest. With respect to the primary purpose of the present study, the findings presented here also make it clear that the distractor-free test of recognition can provide an index of false recognition by simply presenting distractors immediately following the targets.

With respect to the secondary purpose of this study, the results indicate that the target-free test applies within each of the three levels of word frequency used. Lack of a direct correspondence with the results of Wallace et al. (1978) is likely the result of different word-frequency norms. Wallace et al. (1978) selected their words from Thorndike and Lorge's (1944) list, a list based on a count of the frequency of occurrence of the words in the written language, whereas the words of the present study were selected from Ley and Tesiny's (1975) list, a list in which the assessment of associative frequency is based on responses

Table 3
Mean Number of False Recognitions Under
Two Tests of Recognition

Word Frequency	Test of Recognition			
	Target-Free		Discrimination	
	Mean	SD	Mean	SD
High	3.26	4.73	4.41	2.98
Medium	5.63	6.56	8.17	8.29
Low	5.94	4.20	9.69	4.84
Grand Mean	4.94		7.42	

of individual subjects. Although both lists may be highly correlated, subtle differences in assessment procedures can have a strong effect on the memorability of words in a variety of experimental paradigms (e.g., Dean & Ley, 1977; Ley, 1968, 1977; Ley & Dean, 1973).

Administration of the treatments to subjects within relatively large groups (conditions that would be expected to favor "dishonest" performance, i.e., response bias) resulted in no differences in the hit rates between the distractor-free and discrimination test for all levels of word frequency. Perhaps more important was the unexpected finding that the conditions of group administration led to fewer false recognitions of low-frequency words for the distractor-free test than for the discrimination test, and a trend that suggested fewer false recognitions for the distractor-free test overall. A response-bias hypothesis would make the opposite prediction.

Besides providing support for a method for measuring false recognitions while measuring distractor-free recognition, the results of the present study provide serial-position analyses that point to the comparability of performance under the distractor-free and discrimination test.

In addition to the primary and secondary purposes of the present study, the results presented here extend the application of distractor-free tests of recognition (1) from individual testing to group testing, (2) from word-frequency assessments based on the occurrence of words in the written language to associative frequency (a'), and (3) from common English language words to nonwords.

REFERENCES

BADDELEY, A. D. (1976). *The psychology of memory*. New York: Basic Books.

CROWDER, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

DEAN, J., & LEY, R. (1977). Effects of associative encoding on free recall in high and low verbal associators. *Journal of Experimental Psychology: Human Learning & Memory*, 3, 316-324.

KLATZKY, R. L. (1975). *Human memory: Structures and processes*. San Francisco: Freeman.

KLATZKY, R. L. (1975). *Human memory: Structures and processes*. San Francisco: Freeman.

LEY, R. (1968). Associative reaction time, meaningfulness, and presentation rate in paired-associate learning. *Journal of Experimental Psychology*, 78, 285-291.

LEY, R. (1977). Encoding specificity and unidirectional associates in cued recall. *Memory & Cognition*, 5, 523-528.

LEY, R., & DEAN, J. (1973). Effects of associative reaction time and meaningfulness in free recall of mixed and unmixed lists. *Journal of Experimental Psychology*, 98, 220-222.

LEY, R., & KARKER, J. (1982). A comparison of three methods of measuring recognition memory in high and low academic achievers. *Bulletin of the Psychonomic Society*, 19, 11-14.

LEY, R., & TESINY, E. (1975). Associative reaction time, meaningfulness, and pronunciability ratings of 382 words and paralogs. *Bulletin of the Psychonomic Society*, 6, 645-648.

MANDLER, G., GOODMAN, G., & WILKES-GIBBS, D. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, 10, 33-42.

THORNDIKE, E. L., & LORGE, I. (1944). *Teachers' word book of thirty thousand words*. New York: Columbia University Teachers' College.

TULVING, E., & WATKINS, M. J. (1973). Continuity between recall and recognition. *American Journal of Psychology*, 86, 739-748.

UNDERWOOD, B. J. (1972). Word recognition memory and frequency information. *Journal of Experimental Psychology*, 94, 276-283.

WALLACE, W. P. (1978). Recognition failure of recallable words and recognizable words. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 441-452.

WALLACE, W. P. (1980). On the use of distractors for testing recognition memory. *Psychological Bulletin*, 88, 696-704.

WALLACE, W. P., SAWYER, T. J., & ROBERTSON, L. C. (1978). Distractors in recall, distractor-free recognition, and the word-frequency effect. *American Journal of Psychology*, 91, 295-304.

WOODWORTH, R. S. (1938). *Experimental psychology*. New York: Holt.