DOCUMENT RESUME

ED 087 809 TM 003 432

AUTHOR	Ley, Ronald; Karker, Jurgen
TITLE	Measures of Short-Term Recognition Memory.
SPONS AGENCY	National Science Foundation, Washington, D.C.; State Univ. of New York Research Foundation, Albany.
PUB DATE	Nov 73
NOTE	13p.; Presented at the Annual Meeting of the Psychonomic Society (St. Louis, Missouri, November, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Ability; *Cognitive Measurement; High School Students; Low Ability Students; *Memory; Recall (Psychological); *Recognition; Verbal Learning

ABSTRACT

The primary purpose of the present study was to compare three measures of recognition memory. One hundred and forty-four high school students were instructed to study silently 36 CVCVC verbal units (target items) at a 4-sec. rate of presentation. The Ss were then given one of three tasks which required them to identify target items from among distractor items (also CVCVC verbal units). The tasks were the Embedded Item, Multiple Choice, and Shepard-Teghtsoonian. Results demonstrated that the Multiple Choice task yielded the greatest number of correct recognitions, lowest number of false recognitions, and lowest number of nonrecognitions when compared with the other two recognition tasks. Although scores on each of the three tasks correlated with meaningfulness (M), associative reaction time (RT), and pronunciability ratings (PR) of the target items, the magnitude of the coefficients for M and RT were almost identical (rs ranged from .41 to .67) but greater than the coefficients between PR and recognition scores (rs ranged from .36 to .59). The magnitude of the coefficients for the assessment values was greatest for the Embedded Item task and least for the Shepard-Teghtsoonian. (Author)

Abstract

Measures of Short-Term Recognition Memory

Ronald Ley and Jurgen Karker

State University of New York at Albany

(Scheduled for presentation at the Annual Meeting of the Psychonomic Society, St. Louis, Missouri, November, 1973.)

The primary purpose of the present study was to compare three measures of recognition memory. One hundred and forty-four high school students were instructed to study silently 36 CVCVC verbal units (target items) at a 4-sec. rate of presentation. The <u>S</u>s were then given one of three tasks which required them to identify target items from among distractor items (also CVCVC verbal units). The tasks were the Embedded Item, Multiple Choice, and Shepard-Teghtsoonian. Results demonstrated that the Multiple Choice task yielded the greatest number of correct recognitions, lowest number of false recognitions, and lowest number of nonrecognitions when compared with the other two recognition tasks.

Although scores on each of the three tasks correlated with meaningfulness (M), associative reaction time (RT), and pronunciability ratings (PR) of the target items, the magnitude of the coefficients for M and RT were almost identical (<u>r</u>s ranged from .41 to .67) but greater than the coefficients between PR and recognition scores (<u>r</u>s ranged from .36 to .59). The magnitude of the coefficients for the assessment values was greatest for the Embedded Item task and least for the Shepard-Teghtsoonian.

Measures of Short-Term Recognition Memory[1]

Ronald Ley and Jurgen Karker

State University of New York at Albany

Current research in retention has focused on recognition and recall as separate memorial processes. While methods of measuring recall are relatively uniform, methods of measuring recognition are less so. The three methods most commonly used to measure recognition memory are: the Embedded Item task (EI), the Multiple Choice task (MC), and the Shepard-Teghtsoonian task (ST). The EI task typically has a format of verbal units appearing in columnwise fashion on one or more test pages. The task requires S to review the relatively large list and identify verbal units (target items) which were presented during an earlier study period, from among those units which were not presented during the study period (distractor items). Distractor items usually outnumber target by a ratio of at least 2:1. The MC task requires S to select the target item from among a relatively small set of verbal units, e.g., one target item per set of two to four distractors. The ST task requires S to sort a deck of cards in which each card contains one verbal unit. Subjects are required to sort target-item cards, where the number of distractor cards to target cards is in the ratio of about 2:1.

The primary purpose of the present study was to compare these three methods of measuring recognition with respect to performance scores (correct recognitions, false recognitions, and nonrecognitions).

The secondary purpose of the present study was to determine whether the three methods of measuring recognition memory interact with academic ability. That is, if recognition memory is a significant factor in academic achievement, high ability students would be expected to make more correct recognitions, fewer false recognitions, and fewer nonrecognitions than low ability students. However, it is conceivable that part of the expected difference in recognition performance between high and low ability students may be a function of the method used for measuring recognition memory. This possibility has important implications for research designed to test theories of recognition memory and research designed to measure academic achievement or intellectual aptitude in terms of recognition memory scores (e.g., multiple choice tests).

The tertiary purpose of the present study was to compare three assessment characteristics (associative reaction time--RT, meaningfulness--M, and pronunciability ratings--PR) of the verbal units used in the study with respect to their relative capabilities for predicting correct recognition scores of the high and low ability groups under each of the three recognition tasks.

Method

Subjects.--The 144 *Ss* were high school seniors, 72 of whom were randomly selected from the top third of their class with respect to grades and 72 of whom were randomly selected from the bottom third of their class. All *Ss* were naive with respect to previous experience in verbal learning studies.

Materials.--Thirty-six target items were randomly selected from the list of 319 CVCVC words and paralogs assessed for associative reaction time (RT) by Taylor and Kimble (1967), for meaningfulness (M--scaled-rated) by Locascio and Ley (1972), and for pronunciability (PR) by Ley and Karker (1973). The 72 distractor items were selected from the remaining 283 units of this list plus the CVCVC units of Andreas' (1972) list. The target and distractor items were randomly selected with the single restriction that none of the items shared formal similarity with respect to the first three letters of the CVCVC units, i.e., no two of the first three letters of a unit appear more than once in the entire list of 108 items (36 targets and 72 distractors).

The materials for the EI task consisted of a single sheet of $8\frac{1}{2}$- by 11-in. paper on which four columns of 27 units each were printed. The column and row location of each target and distractor item was randomly determined.

The materials for the MC task consisted of 48 decks of 36 Hollerith computer cards. Each card contained one target item and two distractor items printed on the top row of the card. The ordinal position of the target and distractor items on each card was randomly determined.

The materials for the ST task consisted of 48 decks of 108 Hollerith computer cards. Each card contained either a single target item or a single distractor item printed on the top row of the card.

The materials used for the presentation of the target items during the study period consisted of 36 35-mm. slides each one of which contained a single target item. The slides were projected on a film screen by means of a Kodak Carousel slide projector. Presentation rate was controlled by the internal timer of the projector.

Procedure.--During the study phase, *S*s were seated facing the film screen. All *S*s received uniform instructions which required that they silently rehearse each unit as it appeared on the screen. The *S*s were given four practice trials for the purpose of familiarizing them with the types of verbal units which were to be studied and the 4-sec. presentation rate.

Following the study phase, Ss were given brief instructions relevant to their respective recognition task, i.e., Ss in the EI Group were required to read down the typed sheet and to circle with a pencil those items presented during the study phase; Ss in the MC Group were required to circle that unit from among the three units printed on each card, which had been presented during the study phase; and Ss in the ST Group were required to sort their deck by placing cards containing items presented during the study phase in one pile and cards containing other items in a second pile. All Ss were strongly discouraged from guessing.

Results

Three performance measures for each S under each of the three tasks were determined on the bases of (a) correct recognitions (total number of target items identified as "old"), (b) false recognitions (total number of distractor items identified as "old"), and (c) nonrecognitions (total number of target items not identified as "old"). The mean numbers of correct recognitions for the high- and low-ability groups under each of their respective recognition task conditions are given in Table 1 ($n = 24$ Ss per cell).

Insert Table 1 about here

Although the observed mean number of correct recognitions was greatest for the MC task and least for the ST task, a 3 X 2 analysis of variance (three tasks by two ability groups) indicated that differences among the tasks were not significant, $F\ (2, 138) = 1.24$, $p > .05$; but differences between the groups favored the high ability Ss, $F\ (1, 138) = 63.09$, $p < .001$. The Task X Ability Group interaction was not significant, $F\ (2, 138) = .13$, $p > .05$.

With respect to false recognition scores, the analysis of variance indicated that the MC task resulted in the least number of false recognitions and the ST in the most; F (1, 138) = 5.34, $p < .01$. Although the observed differences between the low and high ability Ss were relatively small, the low ability Ss made significantly fewer false recognitions than the high ability Ss; F (1, 138) = 8.48, $p < .001$. The Task X Ability Group interaction was not significant; F (2, 138) = 1.09, $p > .05$.

With respect to nonrecognition scores, the analysis of variance indicated that the ST task resulted in the most nonrecognitions and the MC task in the least; F (2, 138) = 5.16, $p < .01$. Although the observed difference in the mean number of nonrecognitions under the MC task favored the low ability Ss, the overall test of the ability factor resulted in significantly fewer nonrecognitions for the high ability Ss; F (1, 138) = 5.82, $p < .01$. Furthermore, the Task X Ability Group interaction was not significant; F (2, 138) = .16, $p > .05$.

Correlation coefficients between the three assessment characteristics of the target items and number of correct recognitions under the three tasks and two ability groups are given in Table 2. Although all of the correlations are significant, several distinct trends are apparent:

Insert Table 2 about here

(a) The differences between the correlation coefficients for the high and low ability groups within each of the recognition tasks are virtually nil--except for the PR by EI cells, where the correlation for the high ability group (r = -.59) is considerably larger than that for the low ability group (r = -.36). (b) The size of the correlation coefficients between the assessment characteristics and correct recognitions are ranked with respect to task, i.e., rs are largest under the EI task (except for the PR by low ability group, rs range from .59 to .67) and smallest under the ST task (rs range from .40 to .53). (c) The correlation coefficients between RT and correct recognitions and M and correct recognitions are virtually the same under the EI and MC tasks and for each of the ability groups within each of these tasks. Under the ST task, M appears to be a better predictor than RT for both ability groups. (d) The correlations between M and correct recognitions and RT and correct recognitions are larger than those between PR and correct recognitions under all three tasks and for both ability groups.

Discussion

The Embedded Item, Multiple Choice, and Shepard-Teghtsoonian recognition tasks provide very similar data when compared on the bases of correct recognitions. As expected, the high ability high school students made a greater number of correct recognitions than the low ability students; and the differences between the ability groups was almost uniform among the three tasks. It should be noted that the variances of these scores were very similar; S.D.s ranged from 4.50 to 5.95.

The results of the analysis of the false recognitions indicated that the ST task yielded the largest number and the MC task the smallest. The exceptionally small number of false recognitions on the MC task coupled with the relatively small S.D.s suggest that the MC task may be subject to a floor effect. If false recognitions are of special interest to the researcher, the ST task would appear to be the most suitable.

The results of the analysis of the differences between the ability groups on false recognitions under each of the three tasks indicated that the low ability Ss made fewer false recognitions than the high ability Ss. This finding was unexpected. Since the Task-by-Ability-Group interaction was not significant it seems safe to conclude that this finding is not an artifact of the type of task. It may be that this difference reflects a more conservative approach to test taking on the parts of the low ability Ss as compared to the high ability Ss. That is, low ability (low scholastic achievement) Ss by definition, have been incorrect on tests more frequently than high ability (high scholastic achievement) Ss. If subsequent information that "one's answer to a test question is wrong" constitutes punishment, then it might be expected that the more frequently punished Ss (low ability Ss) will require greater confidence in the correctness of their response before emiting the response. This hypothesis is partially corroborated by the nonrecognition scores, which indicated that the low ability Ss made more nonrecognitions than the high ability Ss. Although the Task-by-Ability-Group interaction was not significant, it should be noted that the nonrecognition scores under the MC task were in the opposite direction, i.e., the observed mean number of nonrecognitions was greater for the high ability Ss than the low. In this connection, it should also be noted that the significant Task factor on nonrecognitions was due almost entirely to the relatively low number of nonrecognitions under the MC task, especially among the low ability Ss; whereas scores under the EI and ST task were very similar for both groups.

With respect to the primary purpose of the present study, the comparisons of the three recognition tasks suggest that: (a) they are about equally suited for the purpose of measuring correct recognitions, (b) the MC task, because of floor effects, may be unsuitable for measuring false recognitions, whereas the ST task is clearly most sensitive for this purpose, and (c) the MC task produces low nonrecognitions relative to the EI and ST tasks. In view of the floor effects on false recognition scores and the relatively low number of nonrecognitions, the MC task seems least suitable for measuring recognition memory. Although the EI task produces correct recognitions and nonrecognitions equal to the ST task, it should be noted that false recognitions are almost equally low with respect to the MC task. With these considerations in mind, the best overall test of recognition appears to be the ST task.

With respect to the secondary purpose of the present study, none of the data indicated an interaction among the three measures of recognition and academic ability. The only suggestion of an interaction was found in the nonrecognitions, where scores of the low ability Ss were greater than those of the high ability Ss under the EI and ST task, but not the MC task; the observed means under the MC task were in fact in the opposite direction, i.e., scores of high ability Ss were greater than low ability Ss. This inconsistency with respect to the EI and ST tasks provides another reason for not recommending the MC task as a measure of recognition memory.

With respect to the tertiary purpose of the present study, the analysis of the correlations indicated that: (a) meaningfulness (M) and associative reaction time (RT) were equally better predictors of correct recognitions under the EI and MC tasks than pronunciability ratings (PR), but M was a better predictor than both RT and PR under the ST task; (b) M and RT were most highly correlated with correct recognitions under the EI task and least under the ST task; (c) the correlations between PR and correct recognitions were uniformly moderate under all tasks except for the high ability Ss under the EI task ($r = -.59$); (d) differences between the correlation coefficients of the high and low ability Ss within each task were remarkably similar, i.e., except for the PR and EI task correlations ($r = -.59$ and $r = -.36$), the largest difference was between the M by EI correlations ($r = .67$ and $r = .61$).

REFERENCES

Andreas, B. Meaningfulness (M) values, common associates for paralogs and words, of CVCVC format. Human Engineering Laboratory, U. S. Army Tech. Memo. 27-72; November, 1972.

Ley, R. & Karker, J. Pronunciability ratings of 319 CVCVC words and paralogs previously assessed for meaningfulness and associative reaction time. Submitted for publication to Psychonomic Science.

Locascio, D. & Ley, R. Scaled-rated meaningfulness of 319 CVCVC words and paralogs previously assessed for associative reaction time. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 243-250.

Taylor, J. & Kimble, G. Association value of 320 words and paralogs. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 744-752.

FOOTNOTES

1. This research was supported in part by a State University of New York Research Foundation Grant-in-Aid and an Institutional Grant from the National Science Foundation to the first author.

Table 1

Mean Numbers of Correct Recognitions,

False Recognitions, and Nonrecognitions

for the High and Low Ability

Groups Under Each

Recognition Task

Performance Measure	Academic Ability Group	Recognition Task					
		Embedded Item		Multiple Choice		Shepard-Teghtsoonian	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Correct Recognitions	High	27.38	5.52	29.91	5.60	26.42	4.50
	Low	21.54	5.36	25.87	4.71	22.88	5.95
False Recognitions	High	4.71	3.35	3.33	3.06	7.63	6.56
	Low	4.58	4.47	2.66	2.88	6.29	4.60
Non-Recognitions	High	8.63	5.36	5.42	5.98	9.25	4.75
	Low	14.38	5.62	3.88	4.72	12.88	6.04

Table 2

Product-Moment Correlation Coefficients Between

Assessment Measures of the Target Items

And Correct Recognition Scores

Under the Three Recognition Tasks

And Two Ability Groups (df=34)

Assessment Measure	Recognition Task: Embedded Item, High Ability	Embedded Item, Low Ability	Multiple Choice, High Ability	Multiple Choice, Low Ability	Shepard-Teghtsoonian, High Ability	Shepard-Teghtsoonian, Low Ability
Meaningfulness	.67	.61	.52	.55	.48	.53
Associative Reaction Time	-.64	-.62	-.55	-.52	-.41	-.42
Pronunciability Ratings	-.59	-.36	-.44	-.40	-.40	-.40

Note.--r = .33, p = .025; r = .42, p = .005.