

Verbal discrimination learning and retention as a function of task and performance or observation

MELVIN H. MARX, ANDREW L. HOMER, and KATHLEEN MARX
University of Missouri, Columbia, Missouri 65211

Learning and retention of verbal materials were compared in high school students under perform (guessing with feedback) or observe (watching the performer) conditions. The correct response was randomly determined, or the shortest word, or the word that made a short message when linked to contiguous correct words. In acquisition, subjects did better when performing, females did better than males, and the shortest word group did better than the other groups. There was a variety of interactions in the retention tests, for example, that between scoring (observing subjects either scored their paired performers' responses as right or wrong or simply observed their paired performers) and trials. A transitory inhibitory effect could be attributed to the scoring operation, thus partially confirming a result reported earlier.

The primary purpose of this experiment was to compare learning and retention as a function of performance or observation during acquisition of three types of verbal materials varying in meaningfulness of task materials. A secondary purpose was to assess the role of scoring of performers' responses by observers for two of the materials.

Although a variety of differences between performer and observer in learning situations has been reported (e.g., Marx & Witter, 1972; Witter, Marx, & Farby, 1976), learning task has not been systematically manipulated. Also, the inhibitory influence of observer scoring has been demonstrated (Marx, 1979), but not with verbal learning materials.

METHOD

Design

The task variable was manipulated in three groups presented the same verbal discrimination task of 12 four-item word sets, with different correct answers: randomly determined for the rote (R) group, systematically determined by a rule (three-letter words correct) for the concept (C) group, and meaningfully determined in that four successive correct words constituted a meaningful phrase for the message (M) group. Additional rote and concept groups were run without observer scoring (Rns and Cns) for comparison with the observer-scoring condition in the R and C groups.

This work was supported in part by a Research Career Award from the National Institute of Mental Health and a grant from the Army Research Institute for Behavioral and Social Sciences, both to the first author. The opinions herein are those of the authors alone and do not indicate endorsement by the U.S. Army. We thank Martha Dunham, psychology teacher at Hickman High School, Columbia, Missouri, for her cooperation in providing participants for this research, which was conducted in accordance with university regulations concerning rights of participants. We also thank Maureen Findley, Paula Anderson, and Michael Courtois for assistance in analysis of the data.

Subjects

The participants were 126 high school psychology students trained and tested in their normal classroom setting (class sizes of 24, 22, 24, 26, and 30); each class constituted one of the five groups.

Procedure

Two sets of 12 four-word items were presented to the participants, who were arranged in like-sex pairs. Subjects were told that one word in each set of four was correct and that their task was to discover and remember as many of the correct words as possible. Participants alternated between performance and observation. The performer in each pair decided which of the four words to select and marked the record sheet accordingly. The four words were presented by a slide projector set to advance at an 8-sec rate. One experimenter prompted responding after 3 sec, and the other experimenter called out the correct answer after 6 sec. At that time the observers in the normal scoring condition marked their performers' answers as correct (check mark) or incorrect (cross), and those in the no-score condition simply noted the response and the outcome. After all of the 12 slides were completed, the participants exchanged perform/observe roles, and the second set of 12 slides was presented in the same manner.

For Group C, the correct words in both sets were short (all three letters; e.g., boy), and all of the other words were longer (e.g., wood, edge, and stay, to complete the slide). For Group R, the same two sets of slides were used, but correct words were randomly determined. For Group M, correct words from four successive slides produced a meaningful message when combined sequentially (e.g., chop wood light fire; study hard pass course). Following each of the five acquisition trials, a test trial on both lists was administered, without knowledge of results, in order to compare acquisition under performance and observation.

At the conclusion of the last training-test trial, a free recall test was given: All participants were requested to write down as many of the correct words as they could remember within 3 min. A simple open-ended question was also asked. All participants were to indicate how they had gone about learning the correct words.

Long-term retention tests were administered, free recall followed by cued recall (original four-word slides shown), without advance notice 4 months following the training session.

At that time the experimenter also explained the purpose of the experiment, gave some preliminary training results, and discussed and answered questions on psychology as a college major and a career.

RESULTS

Unweighted means ANOVAs were performed on the data. All of the groups were involved, except when the scoring variable was assessed. In those ANOVAs, Group M, which did not have a nonscoring condition, was not included.

Acquisition

The task factor was a reliable main effect [$F(2,120) = 5.30, p < .05$], with Group C averaging 9.0 correct words, Group M, 7.9, and Group R, 7.6. Task interacted with trials [$F(8,480) = 2.82, p < .05$], so that the initial difference (means of 5.7 correct words for Group C, 4.0 for Group M, and 4.6 for Group R) was much reduced by the fifth trial (means of 10.9 for Group C, 10.9 for Group M, and 10.0 for Group R). Reliably more perform than observe words (means of 8.4 and 8.2, respectively) were correct [$F(1,120) = 6.23, p < .05$], although this difference was obviously very small.

A marginal interaction of task and the perform/observe variable was also found [$F(2,120) = 2.56, p = .08$], with perform means of 9.0, 8.2, and 7.8 and observe means of 9.1, 7.5, and 7.4 for Groups C, M, and R, respectively.

Sex of subject was a reliable main effect; the mean number of correct words was 7.4 for males and 8.8 for females [$F(1,20) = 17.96, p < .01$].

Although there was no reliable main effect for the scoring variable, it did interact reliably with trials [$F(4,384) = 4.15, p < .01$]. This interaction reflects early superiority of nonscoring subjects (5.8 compared with 4.4 mean correct words on Trial 1) and the disappearance of this difference by the end of training (comparable means of 10.4 and 10.6 on the last trial).

Analysis of the responses to the open-ended question provided evidence that 70% of Group C and 68% of Group M had "caught on" to the three-letter rule and the messages; however, only 14% in the latter case specified the four-word structure of the messages. (The two judges who scored these answers showed substantial agreement; interjudge correlations were .87 and .86 for Group C and Group M, respectively).

Immediate Retention

ANOVA for all groups indicated that females recalled more correct words than did males [$F(1,120) = 4.60, p < .05$]. Females recalled a mean of 1.2 correct words, males, .8. This difference, however, cannot be unambiguously attributed to retention per se, because there was a reliable female advantage on the last trial of acquisition, 11.1 compared with 9.7 correct responses [$F(1,120) =$

7.30, $p < .01$]. As shown in Figure 1, there was also an interaction between the scoring and task variables [$F(1,96) = 5.01, p < .05$]. Subjects in Group C who did not score recalled more correct words. That this difference can be attributed to retention is indicated by the fact that there was no reliable last-trial task difference [10.9 correct responses for Group C, 10.9 for Group M, and 10.0 for Group R; $F(2,120) < 1.00$].

Some interesting results were obtained in the recall of incorrect words (i.e., words that had been presented but were not correct). These results are plotted in Figures 2 and 3.

Figure 2 indicates that subjects who scored as observers made more erroneous recalls of perform words, in both Group C and Group R, than subjects who did not score as observers. With respect to the observe words, scoring subjects in Group R made more erroneous recalls, whereas nonscoring subjects in Group C made more erroneous recalls. The reliable three-way interaction [$F(1,96) = 4.34, p < .05$] was due primarily to the marked difference between scoring and nonscoring subjects in Group C on observe words.

The interaction of Sex of Subject by Perform/Observe [$F(1,96) = 4.31, p < .05$] for incorrect recalls is evident in Figure 3. Also evident is the three-way interaction of the sex, perform/observe, and score/no-score variables [$F(1,96) = 6.64, p < .05$]. Scoring subjects of both sexes incorrectly recalled more perform than observe words, while nonscoring subjects showed the opposite

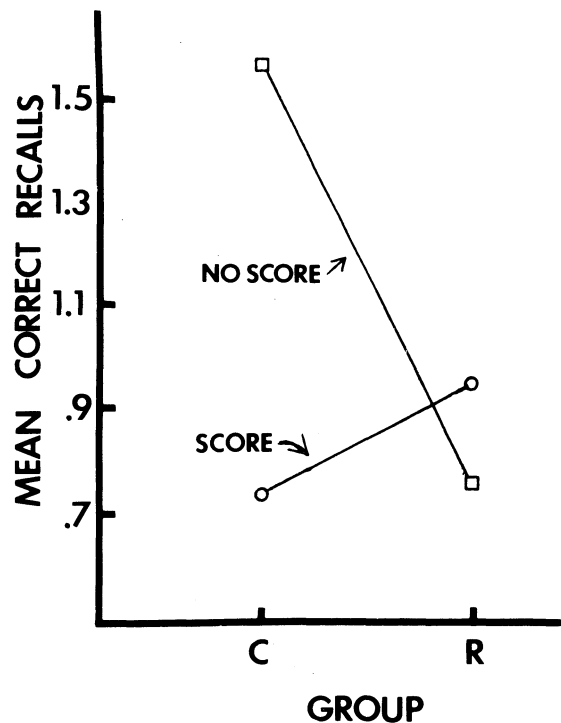


Figure 1. Mean number of correct words recalled in immediate retention as a function of group and scoring variables.

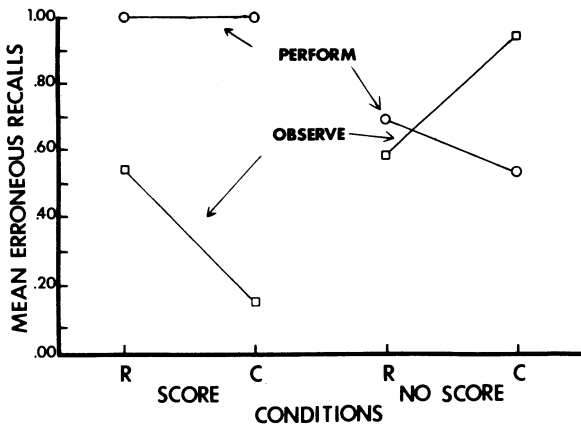


Figure 2. Mean number of words erroneously recalled in immediate retention as a function of group, scoring, and perform/observe variables.

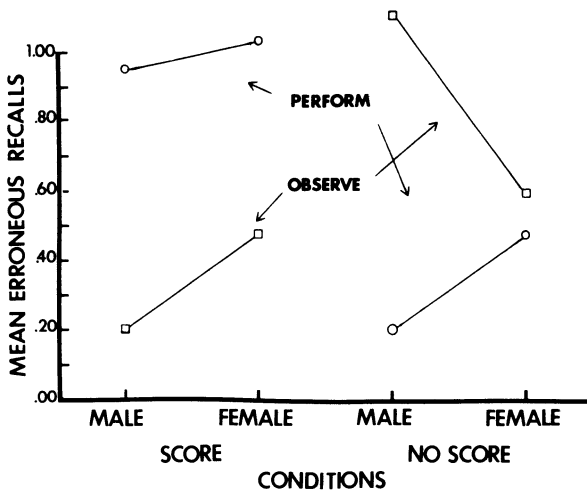


Figure 3. Mean number of words erroneously recalled in immediate retention as a function of sex of subject, scoring, and perform/observe variables.

tendency. The latter trend was especially marked in male subjects.

Long-Term Retention

No statistically reliable effects were found in the free recall test for number of correct words recalled.

A reliable task effect did occur for incorrect responses [F(2,95) = 9.69, p < .01]. The means were .5 for Group R, .2 for Group C, and 1.0 for Group M. There was also a reliable Sex by Perform/Observe interaction [F(1,95) = 3.96, p < .05]; male subjects recalled more incorrect words in the perform condition (means of .6 and .4 for male recall of perform and observe words, respectively, and corresponding female means of .4 and .4).

The task variable provided the only reliable effect in the cued recall test [F(2,95) = 35.95, p < .01].

The mean numbers of correct recalls were 10.7 for Group C, 8.7 for Group M, and 5.4 for Group R.

DISCUSSION

The marginal acquisition interaction between the task and perform/observe variables is mildly suggestive of the relationship postulated in the rationale underlying the design of the experiment; the perform operation, with its presumed more active role of reward, was expected to be more effective in the rote than in the conceptual task. Nevertheless, on the basis of these data, it is safest to note that the performance and observation procedures seem to be essentially equivalent over the task dimension explored.

Introducing a relatively simple rule (only three-letter words correct) clearly facilitated acquisition. Group C also showed the least amount of incorrect long-term recall and a markedly greater amount of long-term cued recall of correct words, indicating that this simple concept was well retained over the 4-month interval. However, these results, like the anticipated acquisition superiority of females on the present verbal materials (cf. Maccoby & Jacklin, 1974), are hardly surprising.

More novel results of greater interest concern the scoring variable and its interactions, mainly with sex of subject and performance/observation, but also to some extent with task. First, it is quite clear that these data provide no simple replication of the generalized inhibitory influence of scoring previously found to be associated, in a task involving the identification of line-drawn faces, with male subjects (Marx, 1979). Nevertheless, the reliable interaction between trials and the scoring variable does provide evidence for at least a transitory general inhibitory influence of scoring. The fact that nonscoring subjects in Group C recalled reliably more correct words in the immediate retention test is also suggestive of interference by the scoring operation. Apparently, the scoring procedure was sufficiently distracting to inhibit both early acquisition and subsequent retention, the latter in spite of equivalent final acquisition levels for scoring and nonscoring groups.

The various reliable interactions found for the scoring variable with the erroneous recall measure also offer material for discussion. The most general conclusion to be drawn from these results is that, apparently, scoring permitted superior tagging of incorrect words, so that words that subjects observed and scored were much less often recalled erroneously than words that they selected as performers. This result occurred both for Groups C and R (cf. Figure 2) and for males and females (cf. Figure 3). This conclusion is supported by the fact that Group C was particularly effective in not recalling very many scored words erroneously, even while erroneously recalling a relatively high number of observed but not scored words (cf. Figure 2). A similarly marked interaction of this type also occurred for male subjects (cf. Figure 3).

We are thus confronted with the seemingly contradictory facts that scoring (1) interferes with both acquisition and retention but (2) helps the subject identify incorrect words properly and so reduces erroneous recall. This apparent contradiction may be resolved if we first assume that observers tend to focus on correct responses more than performers, as has been shown for two-choice discrimination tasks at least (Eelen & d'Ydewalle, 1976). Introduction of the scoring operation, then, presumably interferes with this process by forcing the observer to pay more attention to (and, presumably, to encode more firmly) errors, as well as correct responses. Thus there is some reduction in learning efficiency, relative to the nonscoring condition at least, but there is superior identification, and hence less subsequent recall, of errors.

The clear-cut difference in long-term cued recall, with the Group C mean score approximately twice as large as the

Group R mean score and Group M intermediate, is readily interpretable. The simple concept, identified by at least 70% of Group C, was apparently quite well recalled, so that only three-letter words were selected. The fact that few of the words themselves were correctly identified in the long-term free recall is, of course, consistent with this proposition. (Unfortunately, without the benefit of hindsight, the long-term retention test was not designed to include an appropriate recognition test, which would have consisted of the original three-letter words mixed in with three-letter distractors, to supplement the more difficult free recall test.) The very low error score for Group C in free recall also suggests the long-term retention of the concept. Finally, the fact that even after 4 months male subjects erroneously recalled relatively more perform than observe words extends the earlier reported tendency for males to repeat higher proportions of their own (perform) responses (Marx, Witter, & Farbry, 1973).

REFERENCES

- ELEN, P., & D'YDEWALLE, G. Producing or observing response-outcome contingencies in a two response alternative task. *Psychologica Belgica*, 1976, **16**, 61-71.
- MACCOBY, E. E., & JACKLIN, C. N. *The psychology of sex differences*. Stanford, Calif: Stanford University Press, 1974.
- MARX, M. H. Multiple-choice learning of line-drawn facial features: I. Inhibitory effects of observer scoring. *Bulletin of the Psychonomic Society*, 1979, **14**, 437-438.
- MARX, M. H., & WITTER, D. E. Repetition of correct responses and errors as a function of performance with reward or information. *Journal of Experimental Psychology*, 1972, **92**, 53-58.
- MARX, M. H., WITTER, D. W., & FARBRY, J. Greater repetition of errors under performance compared to observation in multiple-choice human learning. *Perceptual and Motor Skills*, 1973, **37**, 949-950.
- WITTER, D. W., MARX, M. H., & FARBRY, J. Long-term persistence of response-repetition tendencies based on performance or observation. *Bulletin of the Psychonomic Society*, 1976, **8**, 65-67.

ELEN, P., & D'YDEWALLE, G. Producing or observing response-

(Received for publication December 14, 1979.)