

The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners?

WILLIAM L. CULL and EUGENE B. ZECHMEISTER
Loyola University, Chicago, Illinois

College students' ability to judge whether a studied item had been learned well enough to be recalled on a later test was examined in three experiments with self-paced learning procedures. Generally, these learners compensated for item difficulty when allocating study time, studying hard items longer than easy items, but they still recalled more easy items than hard items and tended to drop items out too soon. When provided with test opportunities during study or a delay between study and judgment, learners compensated significantly more for item difficulty and recalled substantially more. Paradoxically, good and poor learners compensated similarly for item difficulty and benefited similarly from testing during study and from delayed decision making. Thus, although the ability to make metamemory decisions was shown to be important for effective learning, these decisions were made equally well by good and poor associative learners. An analysis of tasks used to investigate metamemory-memory relationships in adult learning may provide an account for this apparent learning ability paradox.

The relationship between individuals' knowledge of their own memory, or metamemory, and their own memory performance has been the focus of much research over the last couple of decades (see, e.g., Maki & Swett, 1987; Nelson & Narens, 1990; Zechmeister & Nyberg, 1982). An important part of this research effort has been that which deals with adult learners' ability to make metamemory decisions about discrete verbal items, such as word pairs presented for associative learning (e.g., Arbuckle & Cuddy, 1969) or individual words presented for free recall (e.g., Zechmeister & Shaughnessy, 1980). Research findings show that adult learners are able to assess fairly accurately the relative difficulty of learning items within a set of verbal items (ease of learning, or EOL, judgment; see, e.g., Underwood, 1966), and that they are also capable of judging whether a presently studied item has been learned sufficiently well to be recalled on a later test (judgment of learning, or JOL; see, e.g., Arbuckle & Cuddy, 1969). With regard to JOLs, it has also been demonstrated that accuracy of metamemory judgments improves if (1) test opportunities are present during study (King, Zechmeister, & Shaughnessy, 1980); (2) study

consists of multiple presentations (Lovelace, 1984); or (3) there is a delay between study and judgment trials (Nelson & Dunlosky, 1991).

An assumption underlying research on metamemory decision making is that such decisions are important for efficient learning. A learner's EOL judgments, for example, are correlated with the amount of time spent studying to-be-learned items on a subsequent study trial (Nelson & Leonesio, 1988). Also, when learners make JOLs after an initial study trial and before an opportunity for additional self-paced study, the amount of time that they spend in studying items when they are given a second study opportunity can be predicted by the learners' earlier memorability ratings (Mazzoni, Cornoldi, & Marchitelli, 1990).¹ Thus, metamemory decisions apparently can help learners allocate study time in a manner that will optimize learning.

Nevertheless, the assumption that optimal learning depends on reliable and effective metamemory decisions is undermined by results of studies showing that good learners do not make better metamemory decisions concerning discrete verbal items than poor learners do. Correlations between EOL predictive accuracy and overall recall (Underwood, 1966), as well as between JOL predictive accuracy and overall recall (Lovelace, 1984; Maki & Swett, 1987), reveal little relationship. Moreover, when differences in learning ability are operationalized on the basis of pretest performance as opposed to performance on the critical learning task, no clear differences in EOL predictive accuracy are apparent between good and poor learners (Kearney & Zechmeister, 1989). Thus, results to date constitute an interesting paradox: metamemory de-

We wish to thank Jim Hall, John Shaughnessy, and Tricia Tenpenny for many comments and suggestions that helped shape the preparation of the final draft of this article. These experiments were conducted as part of a master's thesis submitted by the first author to the Graduate School, Loyola University, Chicago. Requests for reprints should be sent to E. B. Zechmeister, Loyola University, Department of Psychology, 6525 N. Sheridan Rd., Chicago, IL 60626.

—Accepted by previous editor, Margaret Jean Intons-Peterson

cisions appear to be related to efficient learning, but the most efficient learners do not appear to make better metamemory decisions than the least efficient learners do.

The purpose of the present set of experiments was to further investigate possible metamemory differences between good and poor adult learners by using a self-paced learning task. Lovelace (1984) has suggested that experimenter-paced tasks may preclude obtaining a positive relationship between metamemory ability and overall learning because subjects' metamemory decisions are less likely to directly affect study activities and hence influence memory performance. Self-paced learning also has the advantage of having greater ecological validity than experimenter-controlled learning does. When studying vocabulary items, for instance, many learners will proceed at their own pace through a set of items until they decide to halt study.

In the present experiments, we asked subjects to study a set of to-be-learned items until they felt they had learned each item. Multiple study opportunities were made available, and control of these study occasions was given to the learners. Specifically, during study, subjects were asked to assess whether a given item had been learned sufficiently well for them to recall it on a later test; a subject's judgment determined whether or not that item would continue to be made available for study. By assessing good and poor learners' metamemory decision making in a self-paced learning task, we were able to investigate general metamemory-memory relationships in a more ecologically valid learning situation, as well as to provide a more sensitive test of possible metamemory differences between good and poor learners.

Experiment 1 focused on good and poor learners' study-time allocation within a multitrial learning situation. In Experiments 2 and 3, we extended this focus by examining two factors previously shown to influence JOLs: test opportunities during study and delay between study and judgment. The results of all three experiments will be discussed in light of what was revealed about metamemory-memory relationships.

EXPERIMENT 1

In several previous experiments (e.g., those of Nelson & Leonesio, 1988, and Mazzoni et al., 1990), the relationship between metamemory judgments and study-time allocation was investigated by asking subjects to provide metamemory ratings of items prior to a single, critical study trial, during which study time for each item was measured. Although these procedures permit a comparison between metamemory ratings and subsequent study times, they have the disadvantage that metamemory judgments and study are separated in time. In the present experiment, students were asked to learn associations between verbal items and to continue study until each item was known. The presentation of the items was under the control of the learner, and thus the dropping of an item from study constituted a judgment that the item was learned sufficiently well so that it could be recalled later.

This type of JOL would seem to occur frequently in everyday learning situations, as, for example, when students must allocate study time among to-be-learned facts in preparation for an examination. We included to-be-learned items that previous research had shown to differ in ease of learning.

We were interested in answering two questions regarding this multitrial learning situation. First, are good learners more sensitive to item difficulty than poor learners? Second, are good learners better able than poor learners to judge when items are learned sufficiently?

Method

Subjects. Forty-one introductory psychology students enrolled at Loyola University of Chicago participated in the experiment. The subjects were tested individually, and each received course credit for participating.

Materials. The subjects were tested with the use of an IBM-compatible computer. Two lists of paired associates were used. List 1 consisted of 20 word pairs of moderate difficulty according to Underwood's (1982) norms that were created based on learning performance. The second list consisted of 18 pairs of high difficulty and 18 pairs of low difficulty (mixed together randomly). All words were five letters in length, and each pair consisted of an uncommon word as the cue and a fairly common word as the response. Sample pairs are *totem-wives*, *lares-black*, and *fugue-fifty*.

Procedure. For List 1 learning, all subjects were given a stack of flash cards, each containing 1 of 20 associative word pairs. The subjects were informed that they would be given a written cued recall test following study and a brief delay. The subjects were given 5 min to study the stack of flash cards; study was self-regulated. A 3-min filler task (math problems) was given following study and before the cued recall test. At the time of the test, all the cues were listed in a random order on an answer sheet, and the subjects were given as much time as they needed to fill in the corresponding response terms. List 1 learning and testing provided a means of discriminating good and poor associative learners as well as a warm-up for the main experimental task.

Following the cued recall test on List 1, each subject was introduced to the main task with the use of a brief computer-presented sample list that comprised three paired associates each consisting of two common nouns. The procedure for second-list learning was identical to that used for the sample list. Specifically, all 36 word pairs were initially presented for 2 sec with the cue appearing above the response; a new random order of pairs was used for each subject. After the familiarization trial, each pair was presented on the screen (one item above the other) for 5 sec, followed by a 3-sec period in which only the following prompt appeared on the screen: "Do you know the word pair? (Yes = terminate study, No = continue study)." The subjects were instructed that the question was asking them to monitor their present learning state, and to decide whether they had learned the item sufficiently well to recall it on a later test. If subjects responded "yes" by striking a specified computer key, the word pair just presented was dropped from the study list. If subjects responded "no," the word pair was retained for further study. The next word pair was presented immediately following the subject's response. If a response was not made during the 3-sec decision period, the word pair was kept in the list for additional study. Remaining word pairs were presented in a new random order following each study/decision cycle. Word pairs continued to be presented for study until each pair had been dropped from the study list. At that time, subjects were given a 3-min filler task (math problems), followed by a written cued recall test similar to that used for List 1. The subjects were given as much time as they needed to complete the test.

Table 1
Mean Study Trials and Proportion Recalled (Experiment 1)

Measure	Learner							
	Good				Poor			
	Easy Items		Hard Items		Easy Items		Hard Items	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Study trials	2.55	.78	3.21	1.2	3.25	1.3	3.87	1.6
Proportion recalled	.37	.12	.30	.14	.23	.13	.16	.15

Note—*SD*, standard deviation.

Results and Discussion

Good and poor learners were distinguished on the basis of List 1 recall via a median split. Mean proportion recall was .83 for good learners and .33 for poor learners. One subject was randomly dropped to attain an equal number ($n = 20$) of subjects in each group. A 2 (good vs. poor learner) \times 2 (easy vs. hard item) mixed design analysis of variance (ANOVA) was used to analyze the number of study presentations allocated for List 2 as well as the amount recalled. An alpha level of .05 was used for all tests.

The mean number of study trials and mean proportion recall for List 2 learning are summarized in Table 1. Subjects compensated for item difficulty by allocating significantly more study time to hard items than to easy items [$F(1,38) = 53.09$, $MS_e = .16$]. Also, poor learners studied for more study trials than good learners did (see Table 1), although the difference was not significant [$F(1,38) = 3.03$, $MS_e = 3.10$, $p = .09$]. There was no interaction between item difficulty and learning ability for amount of study [$F(1,38) < 1$, $MS_e = .16$]. Thus, good and poor learners compensated similarly for item difficulty; both groups allocated more study presentations to hard items.²

Easy items were recalled better than hard items [$F(1,38) = 43.49$, $MS_e = 2.70$], and good learners recalled more than poor learners [$F(1,38) = 11.35$, $MS_e = 45.37$]. There was no interaction between item difficulty and learning ability for items recalled [$F(1,38) < 1$, $MS_e = 2.70$]. Inspection of the means in Table 1 reveals that neither good nor poor learners approached perfect recall; mean proportion recall ranged from .16 to .37 across the two categories of learners and the two types of items.

One major finding of Experiment 1 is consistent with that of previous research (e.g., Nelson & Leonesio, 1988; Zacks, 1969): subjects study hard items more than easy items, but they recall easy items better than hard items. The present experiment extends this finding to a situation in which multiple study trials are available. The results of Experiment 1 also highlight the paradox regarding the relationship between learning ability and metamemory decision making. Learners who differ widely in initial learning ability apparently compensated equally for differences in item difficulty. This finding is similar to that reported by Lovelace (1984) and by Kearney and Zechmeister (1989), who found little or no correlation between learning ability and JOL and EOL accuracy, respectively. Although

good and poor learners exhibited similar compensatory study, neither group, however, compensated sufficiently. Easy items were still recalled more than hard items, and even in the best condition (good learners studying easy items), only about one third of the word pairs were successfully recalled.

EXPERIMENT 2

In Experiment 2, we further examined good and poor learners' allocation of study time within a multitrial learning situation. Unlike in Experiment 1, however, we included a second experimental condition that provided test trials during study. Alternating study and test trials substantially improves the predictive accuracy of JOLs (King et al., 1980; Shaughnessy & Zechmeister, 1992). We reasoned that providing test trials during study would similarly aid subjects' allocation of study time and improve recall. Of particular interest was whether good and poor learners would benefit differentially from testing. That is, are good learners more likely than poor learners to use information generated from test opportunities or more capable than poor learners of using that information effectively?

Method

Subjects. Sixty-four undergraduates participated. No subject had participated in the first experiment, and all subjects were tested individually and received course credit for their participation in the experiment.

Materials and Procedure. The materials and procedures were essentially the same as in Experiment 1. There were two notable changes, however. First, unlike in Experiment 1, subjects were told (1) that they would have to remain for the entire experimental period, regardless of when the learning task was completed, and (2) that they were being asked to attempt a very difficult task that would require their full attention. These minor changes in instructions were made in order to increase the subjects' motivation to learn the task materials completely.

A second change was the addition of a second presentation condition that was manipulated between subjects: the test-study condition. In this condition, following the familiarization trial, the first word of each critical word pair was presented alone for 2.5 sec, followed by a 2.5-sec period during which both words of the pair were present. Thus, each presentation was 5 sec long, but divided into two 2.5-sec periods. Following each test-study presentation, the subjects were prompted as to whether they knew the word pair. The prompt was identical to that in Experiment 1, and was presented for a maximum of 3 sec. Immediately following a response,

Table 2
Mean Study Trials and Proportion Recalled (Experiment 2)

Measure	Learner							
	Good				Poor			
	Easy Items		Hard Items		Easy Items		Hard Items	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Study-Only Condition								
Study trials	2.82	1.1	3.48	1.3	3.31	2.2	3.89	2.3
Proportion recalled	.72	.30	.59	.33	.46	.25	.28	.20
Test-Study Condition								
Study trials	3.61	.89	4.50	1.3	4.28	1.6	5.57	2.1
Proportion recalled	.85	.17	.80	.21	.58	.30	.46	.28

Note—*SD*, standard deviation.

the next word pair was presented. As in Experiment 1, the word pairs continued to be presented until the subject responded "yes" for each word pair. The subjects were then given a 3-min distractor task (math problems) followed by a written cued recall test.

Thus, there were two presentation conditions: test-study and study-only. The study-only condition was identical to Experiment 1 except for the instructional changes, and it differed from the test-study condition only in the method of presenting items for study.

Results and Discussion

The allocation of study during the second learning task, as well as the number of items recalled, were analyzed in a 2 (good vs. poor learner) \times 2 (test-study vs. study-only) \times 2 (easy vs. hard item) mixed ANOVA design. Good and poor learners were distinguished on the basis of two separate median splits of List 1 recall, one for the study-only condition and one for the test-study condition. Mean proportion recalled for good learners in the study-only condition was .82, and it was .73 in the test-study condition; mean proportions for poor learners were .30 and .23, respectively.

The mean number of study trials and mean proportion recall for List 2 learning are summarized in Table 2. Overall, hard items were allocated more study presentations than were easy items [$F(1,60) = 128.36$, $MS_e = .18$], and subjects in the test-study condition allocated more study presentations than did subjects in the study-only condition [$F(1,60) = 7.13$, $MS_e = 5.52$]. Although poor learners allocated more study presentations than did good learners, the difference was not significant [$F(1,60) = 2.53$, $MS_e = 5.52$] (see Table 2). The three-way interaction of learning ability with item difficulty and testing was not significant [$F(1,60) = 2.45$, $MS_e = .18$]. Likewise, the two-way interactions of learning ability with item difficulty [$F(1,60) = 1.31$, $MS_e = .18$] and of learning ability with testing [$F(1,60) < 1$, $MS_e = 5.52$] were not significant.

Testing did interact with item difficulty [$F(1,60) = 9.29$, $MS_e = .18$]. Figure 1 reveals that hard items were studied longer than easy items in both the test-study condition [$F(1,31) = 73.41$, $MS_e = .23$] and the study-only condition [$F(1,31) = 52.83$, $MS_e = .12$]. The magnitude of this difference, however, was greater in the test-study condition. In other words, subjects exhibited greater com-

penation for item difficulty when provided feedback through testing.

Easy items were recalled more than hard items [$F(1,60) = 60.34$, $MS_e = 2.54$], subjects in the test-study condition recalled more than subjects in the study-only condition [$F(1,60) = 6.15$, $MS_e = 41.15$], and good learners recalled more than poor learners [$F(1,60) = 21.9$, $MS_e = 41.15$]. The three-way interaction of item difficulty, testing, and learner was not significant [$F(1,60) < 1$, $MS_e = 2.54$], and the two-way interaction of learner with testing was not significant [$F(1,60) < 1$, $MS_e = 41.15$]. The two-way interaction of learner with item difficulty also was not significant, although it approached significance [$F(1,60) = 3.15$, $MS_e = 2.54$, $p = .081$]. There was a somewhat greater difference between recall of hard items and recall of easy items for poor learners than for good learners.

The interaction of testing with item difficulty was significant [$F(1,60) = 4.93$, $MS_e = 2.54$]. Although the difference between hard and easy items was greater in the test-study condition when study was analyzed, the reverse was true when recall was analyzed. Simple effect analyses

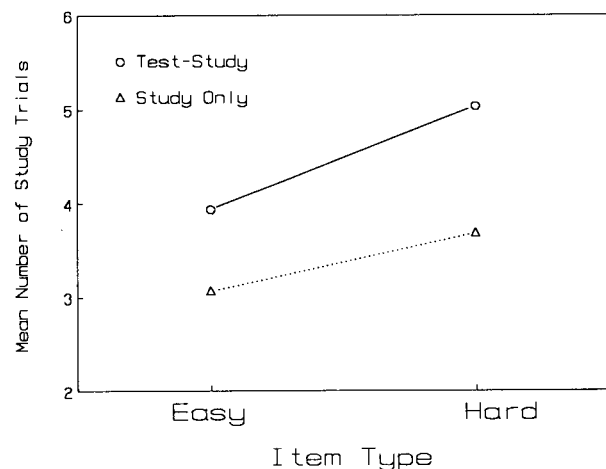


Figure 1. Mean study trials as a function of item type and testing.

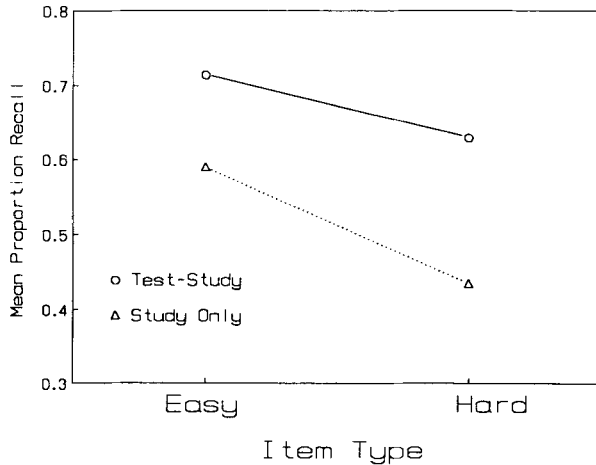


Figure 2. Mean proportion recall as a function of item type and testing.

revealed that easy items were recalled significantly more than hard items both in the test-study condition [$F(1,31) = 19.5$, $MS_e = 2.00$] and in the study-only condition [$F(1,31) = 39.90$, $MS_e = 3.17$]. As can be seen in Figure 2, the difference was greater in the study-only condition.

The results of Experiment 2 replicated and extended those of Experiment 1. The results of Experiment 2 further reveal that learners are sensitive to differences between hard and easy items and compensate for these differences during study. The presence of testing during study appeared to enhance metamemory decision making and to improve overall recall. This is strongly suggested by the complementary interactions between item difficulty and testing for study and recall. Learners compensated for item difficulty during study more in the test-study condition than in the study-only condition, and the difference between recall of hard items and recall of easy items was reduced in the test-study condition. It would appear that test trials aid learners' decisions regarding when to or not to terminate study, just as they improve decisions about what will be remembered (i.e., JOLs).

No significant differences emerged in the ways in which good and poor learners allocated study time. As in Experiment 1, poor learners tended to study slightly (but not significantly) longer than good learners, yet good learners recalled significantly more. Most importantly, good and poor learners were influenced similarly by testing. Both good and poor learners compensated more for item difficulty if provided with testing during study. Also, for both good and poor learners, there was less difference between recall of hard items and recall of easy items with testing. These results replicate and extend the paradoxical relationship found in Experiment 1: Despite considerable differences in initial and final recall, good and poor learners appear equally sensitive to item difficulty and benefit equally from testing during study.

EXPERIMENT 3

Nelson and Dunlosky (1991) have demonstrated that JOL predictive accuracy is significantly better when there is a brief delay between study and judgment trials than when judgments are made immediately following study, at least when only the first member of a to-be-learned pair is presented at the time of judgment; the effect did not occur when both members of the pair were presented (Dunlosky & Nelson, 1992). One explanation of these results is that the cue-only delayed JOL mimics a cued recall test trial and allows learners to base their judgments on the outcome of attempted retrieval (see Spellman & Bjork, 1992; but see also Nelson & Dunlosky, 1992). In other words, subjects use information gained on a "cued recall judgment" trial to assess future recallability of items, just as subjects have been shown to use their performance on test trials during study to make JOLs (e.g., King et al., 1980).

In Experiment 3, subjects were presented with only the first member of the word pairs when asked to make their judgments to continue study or not; judgments were made either immediately following study or after a brief delay. The delay was expected to have an effect similar to that demonstrated for test trials in Experiment 2. Thus, Experiment 3 represented a convergent validity check for the results of Experiment 2, and it provided another opportunity for possible differences between good and poor learners to emerge.

Method

Subjects. Eighty introductory psychology students from the same subject pool as that used in Experiments 1 and 2 participated in this experiment. All subjects were tested individually, received course credit in exchange for their participation, and did not participate in either of the first two experiments.

Materials and Procedure. The materials and procedure were nearly identical to those of the first two experiments. Upon arrival, the subjects were randomly assigned to one of two conditions: immediate or delay. In order to provide a better comparison group for the delay condition, two minor changes were made in the presentation used previously for the study-only condition. First, when the subjects were prompted as to whether they knew a given word pair, the first word of the word pair remained on the screen. Second, the subjects were given a maximum of 8 sec as opposed to 3 sec to decide whether they knew the word pair.

The only difference between the immediate condition and the delay condition was the timing of the presentation of the decision period. In the delay condition, the subjects were not prompted as to whether they knew each pair until after they had studied all the pairs. That is, the subjects were first presented each of the word pairs for 5 sec and were then presented the first word of each pair along with the metamemory prompt. The decision presentations were administered in the same order as were the study presentations, and the subjects had a maximum of 8 sec in which to decide. Word pairs not dropped out were re-presented in a new random order; study presentations always preceded the decision presentations.

Results and Discussion

As before, two main dependent variables, number of study presentations allocated and number of words re-

Table 3
Mean Study Trials and Proportion Recalled (Experiment 3)

Measure	Learner							
	Good				Poor			
	Easy Items		Hard Items		Easy Items		Hard Items	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Immediate Condition								
Study trials	3.10	1.3	3.69	1.8	3.35	1.4	4.09	1.6
Proportion recalled	.74	.22	.61	.30	.39	.26	.26	.22
Delay Condition								
Study trials	2.75	.97	3.65	1.4	3.47	1.2	4.80	1.8
Proportion recalled	.88	.12	.80	.16	.60	.20	.43	.18

Note—*SD*, standard deviation.

called on the final cued recall test, were analyzed. A 2 (good vs. poor learner) \times 2 (immediate vs. delay) \times 2 (easy vs. hard item) mixed ANOVA design was used to analyze each dependent variable. Again, good and poor learners were distinguished on the basis of the first cued recall test. Separate median splits were performed for both conditions. Mean proportions were .74 (good learner) and .34 (poor learner) in the delay condition, and .73 (good learner) and .29 (poor learner) in the immediate condition.

The mean number of study trials and mean proportion recall for List 2 learning are summarized in Table 3. Analysis of the number of study presentations allocated revealed a significant main effect of item difficulty [$F(1,76) = 143.13$, $MS_e = .22$], but no significant main effects of delay [$F(1,76) < 1$, $MS_e = 4.05$] or learner [$F(1,76) = 3.9$, $MS_e = 4.05$, $p = .052$]. In similarity with the findings of Experiments 1 and 2, hard items were studied longer than easy items, and poor learners studied longer than good learners, although this latter difference was again not significant (see Table 3).

The three-way interaction of item difficulty, delay, and learner was not significant [$F(1,76) < 1$, $MS_e = .22$]. Also, the two-way interactions of learner and delay [$F(1,76) < 1$, $MS_e = 4.05$], as well as learner and item difficulty [$F(1,76) = 3.65$, $MS_e = .22$], were not significant, although the interaction of learner and item difficulty approached significance ($p = .06$). The difference between easy and hard items was generally greater for poor than for good learners. The interaction of item difficulty and delay was significant [$F(1,76) = 9.07$, $MS_e = .22$]. Simple effect analyses revealed that hard items were studied significantly more than easy items in both the delay condition [$F(1,39) = 91.63$, $MS_e = .27$] and the immediate condition [$F(1,39) = 47.6$, $MS_e = .19$], with the difference being greater in the delay condition (see Figure 3).

Examination of the number of items recalled on the cued recall test revealed that easy items were recalled significantly more than hard items [$F(1,76) = 80.00$, $MS_e = 2.95$], good learners recalled significantly more items overall than did poor learners [$F(1,76) = 54.97$, $MS_e = 26.5$], and subjects in the delay condition recalled significantly more items than did subjects in the immediate condition [$F(1,76) = 14.8$, $MS_e = 26.5$]. There was no

interaction between learner, item difficulty, and delay [$F(1,76) = 2.9$, $MS_e = 2.95$]. There were also no significant interactions between learner and delay [$F(1,76) < 1$, $MS_e = 26.5$], learner and item difficulty [$F(1,76) = 3.22$, $MS_e = 2.95$], and delay and item difficulty [$F(1,76) < 1$, $MS_e = 2.95$]. It should be noted, however, that the two-way interaction of learner and item difficulty, as well as the three-way interaction of learner, item difficulty, and delay, did approach significance ($ps = .077$ and $.093$, respectively). The disparity between the number of easy items and the number of hard items recalled tended to be less for good learners, and this difference in recall performance by good and poor learners apparently was accentuated in the delay condition.

The results are generally consistent with those of the previous experiments. In addition, Experiment 3 showed that a delay between study and decision led to significantly greater compensation for item difficulty during study. The delay also led to better recall. Once again, there was a notable lack of relationship between learner ability and metamemory decision making, although several interactions with type of learner approached significance. Within each of these interactions, poor learners either

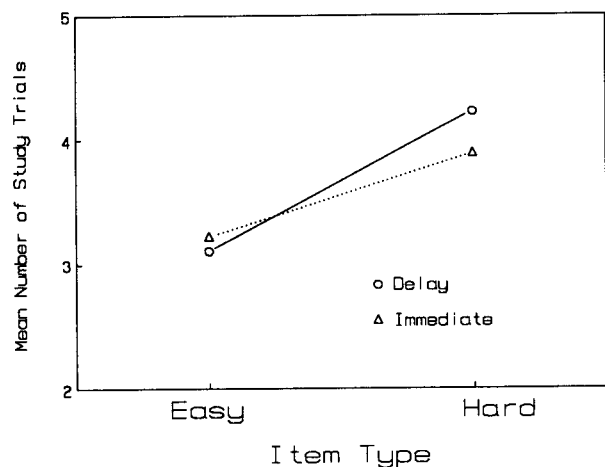


Figure 3. Mean study trials as a function of item type and delay.

studied items longer or compensated more for item difficulty than good learners.

GENERAL DISCUSSION

In the present experiments, we investigated two general kinds of metamemory decision making among young adult learners who differed in associative learning ability. First, by including to-be-learned items that differed substantially in relative ease of learning, we were able to look at the learners' ability to make relative distinctions between items regarding learning difficulty. In addition, by using a self-paced learning task and multiple study opportunities, we were able to examine the learners' ability to make absolute judgments regarding when a given item was sufficiently well learned so that it could be recalled on a later test. The results will be discussed first in terms of general factors shown to affect metamemory decision making and then in terms of what we have learned regarding how good and poor learners make these decisions.

Metamemory Decisions

In the standard self-paced learning condition used in Experiment 1 and in the control conditions of Experiments 2 and 3 (both words presented for 5 sec, followed by a metamemory prompt), the learners demonstrated that they did assess and utilize differences in item difficulty when allocating study, but faltered in their judgments of whether an item had been learned sufficiently. The learners compensated for item difficulty by studying hard items more than easy items, but they clearly dropped items out too soon, on the average. The subjects in the standard learning conditions of Experiments 1, 2, and 3 recalled only .27, .51, and .50 of the items, respectively; the difference in recall between the first of these groups and the second two is likely the result of the instructional changes that were made following Experiment 1. This result pattern is consistent with previous research (Nelson & Leonesio, 1988; see also Mazzoni et al., 1990) reporting that learners halt study prematurely, even when given the opportunity for unlimited study time.

The results for both the test-study condition of Experiment 2 (first word presented for 2.5 sec, and then both words presented for 2.5 sec followed by a 3-sec prompt) and the delay condition of Experiment 3 (both words presented for 5 sec and then a delay and the first word presented during an 8-sec prompt) were dramatically different from those obtained in the standard self-paced learning conditions. In Experiment 2, the subjects in the test-study condition studied items more overall, compensated more for item difficulty, and recalled more than did the subjects in the study-only condition. In Experiment 3, the subjects in the delayed test condition did not study items more overall, but did compensate more for item difficulty and did recall more than did subjects in the immediate test condition. In other words, study under either test-study or delayed JOL procedures led to greater overall recall and greater compensation for item difficulty.

It is important to note that both procedures share a delayed test component but that the consequences of that test for learners can be different. In the test-study procedure of Experiment 2, a learner is always tested on the critical pair immediately before a study opportunity. Given that the critical pair is last seen on the initial familiarization trial or on an earlier test-study trial, the test necessarily is delayed. However, because both members of the to-be-learned pair are available immediately after the test, retrieval on the test need not be successful for the subject to have additional opportunity for study on that trial. In the delayed condition of Experiment 3, although testing is similarly delayed, there is no immediate opportunity for further study after the test unless retrieval is successful. If a retrieval attempt fails, and if the learner decides that more study is desired, another study and test cycle must be initiated.

It seems likely that the presence of retrieval opportunities during study or delayed decision making benefits both metamemory decision making and recall. There is clear evidence that test trials improve the accuracy of metamemory decisions (see, e.g., King et al., 1980; Shaughnessy & Zechmeister, 1992). The "delayed JOL effect" investigated by Nelson and Dunlosky (1991, 1992) also attests to the fact that postponing a prediction concerning an item's future recallability leads to greater metamemory accuracy than does making an immediate prediction. Information gained from attempted retrieval of the target item during the delayed cued recall judgment trial is likely the major source of information for this metamemory decision (see, e.g., Spellman & Bjork, 1992). There is also evidence that successful retrieval of an item after a delay contributes more to an item's memorability than does additional study (see, e.g., Carrier & Pashler, 1992). Thus, learning procedures that incorporate a test-study technique or delayed cued recall trial (or perhaps some combination), are likely to lead to more effective learning than are those involving study-only or study followed by an immediate cued recall procedure.

Metamemory and Learning Ability

Learners who differ widely in initial learning ability compensated equally for differences in item difficulty in a self-paced associative learning task (Experiment 1), and similarly compensated more for item difficulty when provided test opportunities during study (Experiment 2) or a delay between study and judgment (Experiment 3). Moreover, in terms of recall performance, good and poor learners benefited similarly from the presence of test trials during learning or from the addition of a delay between study and a decision to terminate study.

A consistent finding across all three experiments was that poor learners studied critical items as many times, or even more times, as did good learners, but recalled less. Moreover, because the level of recall achieved by good learners more closely approximated the goal given to the learner (i.e., successful recall on a later retention test), it might be suggested that, in terms of absolute

metamemory judgments, good learners performed better than poor learners. Differences in final levels of recall obtained for good and poor learners, however, are likely to have more to do with learning effectiveness than with metamemory. Thus it is unclear, given the present materials and task, to what extent differences in overestimation are due to inherent differences in learning ability or to differences in metamemory ability. What is clear from the present results is that both good and poor learners terminate study too soon on the average.

Why do students who clearly differ in associative learning ability seemingly not differ in their sensitivity to item difficulty, in terms of an effect of testing on recall, or in terms of an effect of delayed metamemory decision making? Why, in other words, do results of the present experiments, as well as results of numerous other experiments (e.g., those of Glenberg & Epstein, 1985; Kearney & Zechmeister, 1989; Lovelace, 1984; Maki & Swett, 1987; Underwood, 1966) not reveal more substantial relationships between learning ability and metamemory decision making? One answer, of course, is that such relationships are nonexistent. Although the ability to make metamemory decisions is important for effective learning, these decisions are made equally well by both good and poor associative learners. Such a position is contradicted, however, by a number of other studies indicating that a relationship between metacognition and learning ability does exist. These studies show, for example, that adults make better metacognitive decisions than do children (e.g., Pressley, Levin, & Ghatala, 1984), that more successful fifth graders monitor learning difficulty of prose passages better than do less successful fifth graders (e.g., Owings, Petersen, Bransford, Morris, & Stein, 1980), that high-achieving college students more accurately discriminate what is known and not known than do low-achieving students (e.g., Shaughnessy, 1979; Zechmeister, Rusch, & Markell, 1986), that college students who perform better on a text comprehension task monitor their comprehension better than students performing less well (e.g., Maki & Berry, 1984), or even that expert memorizers regulate study more effectively than do novice memorizers (see Intons-Peterson & Smyth, 1987).

The resolution to this apparent paradox would seem to lie in an analysis of the tasks used in many metamemory experiments. Studies that have revealed a relationship between cognitive monitoring and learning ability generally have used tasks and materials that are more complex than those used in studies not revealing a relationship. For example, Maki and Berry (1984), as well as Maki and Swett (1987), had subjects read passages of narrative text and predict future test performance. The subjects in the first study made predictions for complete sections of narrative text, whereas the subjects in the second study made predictions for specific idea units. Maki and Berry (1984) found a relationship between task performance and predictive accuracy; Maki and Swett (1987) did not.

Without other information (e.g., prior test trials) a major source of information used to predict the outcome

of a future test of memory is likely to be the ease of processing at time of judgment (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). This in turn will be related to characteristics of the to-be-judged items, such as their meaningfulness, concreteness, pronunciability, and so forth. The word pairs used in the present experiments may be expected to differ in how easily they may be processed. It is possible that sensitivity to ease of processing is a fundamental cognitive process (cf. Zechmeister & Bennett, 1991), which is likely to be present to the same degree for both types of learners in the present JOL tasks, as has been previously demonstrated for an EOL task (see, e.g., Kearney & Zechmeister, 1989). Thus, both types of learners may make similar, albeit sometimes premature, relative judgments to terminate study.

The results obtained in Experiments 2 and 3 can be viewed optimistically in that when discrete verbal items are used, adults of widely different learning abilities benefit similarly from procedures shown to enhance metamemory decision accuracy and recall. Because many school tasks involve memorization of simple facts—for example, in the form of learning scientific terms, new vocabulary, or geographical locations—the present findings indicate that these procedures may be used to aid recall of learners regardless of learning ability. Because a delayed test component is present in both a test-study procedure and a delayed cued recall judgment procedure, there would appear to be clear, concrete evidence given to a learner regarding the success of encoding efforts directed toward target items. This kind of evidence seems to be understood by both good and poor associative learners. However, because poor learners study as much as or more than good learners but remember less, and because all learners generally study hard items more than easy items but recall them less, feedback about an item's memorability is not likely to be sufficient to guarantee successful retention. Learners must be trained in more effective encoding strategies as well. It remains to be seen whether procedures that incorporate delayed testing can, when coupled with training in effective learning strategies, reduce the gap between good and poor associative learners and between memory for easy and memory for hard items.

REFERENCES

- ARBUCKLE, T. Y., & CUDDY, L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*, 126-131.
- BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, *28*, 610-632.
- CARRIER, M., & PASHLER, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633-642.
- DUNLOSKY, J., & NELSON, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374-380.
- GLENBERG, A. M., & EPSTEIN, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 702-718.

- INTONS-PETERSON, M. J., & SMYTH, M. M. (1987). The anatomy of repertory memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 490-500.
- KEARNEY, E. M., & ZECHMEISTER E. B. (1989). Judgments of item difficulty by good and poor associative learners. *American Journal of Psychology*, **102**, 365-383.
- KING, J. F., ZECHMEISTER, E. B., & SHAUGHNESSY, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, **93**, 329-343.
- LOVELACE, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 756-766.
- MAKI, R. H., & BERRY, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 663-679.
- MAKI, R. H., & SWETT, S. (1987). Metamemory for narrative text. *Memory & Cognition*, **15**, 72-83.
- MAZZONI, G., CORNOLDI, C., & MARCHITELLI, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, **18**, 196-204.
- NELSON, T. O., & DUNLOSKY, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The delayed JOL effect. *Psychological Science*, **2**, 267-270.
- NELSON, T. O., & DUNLOSKY, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, **3**, 317-318.
- NELSON, T. O., & LEONESIO, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 676-686.
- NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Vol. 26. Advances in research and theory* (pp. 125-173). New York: Academic Press.
- OWINGS, R. A., PETERSEN, G. A., BRANSFORD, J. D., MORRIS, C. D., & STEIN, B. S. (1980). Spontaneous monitoring and regulation of learning: A comparison of successful and less successful fifth graders. *Journal of Educational Psychology*, **72**, 250-256.
- PRESSLEY, M., LEVIN, J. R., & GHATALA, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning & Verbal Behavior*, **23**, 270-288.
- SHAUGHNESSY, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, **13**, 505-514.
- SHAUGHNESSY, J. J., & ZECHMEISTER, E. B. (1992). Memory-monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society*, **30**, 125-128.
- SPELLMAN, B. A., & BJORK, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, **3**, 315-316.
- UNDERWOOD, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, **71**, 673-679.
- UNDERWOOD, B. J. (1982). Paired associate learning: Data on pair difficulty and variables that influence difficulty. *Memory & Cognition*, **10**, 610-617.
- ZACKS, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology*, **82**, 441-447.
- ZECHMEISTER, E. B., & BENNETT, D. J. (1991). How easy is it to judge ease of learning? *Bulletin of the Psychonomic Society*, **29**, 36-38.
- ZECHMEISTER, E. B., & NYBERG, S. E. (1982). *Human memory: An introduction to research and theory*. Pacific Grove, CA: Brooks/Cole.
- ZECHMEISTER, E. B., RUSCH, K. M., & MARKELL, K. A. (1986). Training college students to assess accurately what they know and don't know. *Human Learning*, **5**, 3-19.
- ZECHMEISTER, E. B., & SHAUGHNESSY, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, **15**, 41-44.

NOTES

1. This was apparent when the initial presentation time was relatively brief (5 sec or less). When the initial presentation time was longer (7.5 sec) subjects restudied longest those items about which they were most uncertain (cf. Mazzoni et al., 1990).
2. Additional analyses comparing the number of study trials allocated to recalled and to unrecalled items were completed for each experiment but were not reported because they were very similar to the results found when comparing the number of study trials allocated to hard and easy items. For both good and poor learners, unrecalled items tended to be studied longer than recalled items, just as hard items tended to be studied longer than easy items.

(Manuscript received December 21, 1992;
revision accepted for publication July 3, 1993.)