Allocation of Self-Paced Study Time and the "Labor-in-Vain Effect"

Thomas O. Nelson and R. Jacob Leonesio
University of Washington

This research explored the possibility that a metacognitive control process (namely, the allocation of self-paced study time) might be affected by the output from metacognitive monitoring processes (i.e., ease-of-learning and/or feeling-of-knowing judgments). In three experiments, university undergraduates received instructions that emphasized either accuracy of learning or speed of learning. The major findings were: (a) ease-of-learning judgments and feeling-of-knowing judgments are reliably related to study-time allocation, with more self-paced study time being allocated to the supposedly more difficult items; (b) even when instructed to master every item and when allowed unlimited study time to do so, people terminate study before learning is completed; and (c) large increases in self-paced study time can yield little or no increase in the subsequent likelihood of recall (the "labor-in-vain effect"). Implications are drawn for a model of the interplay between metacognitive monitoring processes and metacognitive control processes.

Metamemory is generally conceptualized as including monitoring components (e.g., ease-of-learning judgments and feeling-of-knowing judgments) and control components (e.g., allocation of study time in a self-paced learning situation) (Zechmeister & Nyberg, 1982). Although it is reasonable to hypothesize that there should be an interplay between these two kinds of components, previous empirical investigations have focussed on only one kind or the other in isolation.

The present research attempted to bring those two kinds of components together to explore the notion that the control components are affected by the output from the monitoring components. In particular, the issue under investigation was whether the allocation of study time in self-paced learning situations is inversely related to the outputs from ease-of-learning (EOL) judgments and/or feeling-of-knowing (FOK) judgments, that is, the person will allocate more study time to items believed to be difficult (low in EOL or FOK) than to items believed to be easy.

To help connect the present research with the background literature, we will review briefly the previous findings related to this issue and then place them within a model that generates the predictions we examined.

Monitoring: EOL and FOK Prior to Experimenter-Paced Study Trials

The first research we know of that examined self-paced study trials was by Bugelski and Rickwood (1963). They reported the overall study time for the entire list (for their comparison with experimenter-paced situations) but did not examine the distribution of study times across different items. Subsequent research showed that people do allocate different amounts of study time to various items in paired-associates learning (Zacks, 1969), in serial recall (Le Ny, 1969), and in free recall (Belmont & Butterfield, 1971; additional experiments are reviewed in Butterfield & Belmont, 1977, p. 283ff). Most pertinent to the present research, the usual conclusion is "strong evidence that subjects in self-paced procedures study difficult pairs longer than easy ones [defined in terms of the likelihood of being correct during test trials, not in terms of EOL]" (Zacks, 1969, p. 446).

Control: Allocation of Study Time During Self-Paced Study Trials

The first research we know of that examined self-paced study trials was by Bugelski and Rickwood (1963). They reported the overall study time for the entire list (for their comparison with experimenter-paced situations) but did not examine the distribution of study times across different items. Subsequent research showed that people do allocate different amounts of study time to various items in paired-associates learning (Zacks, 1969), in serial recall (Le Ny, 1969), and in free recall (Belmont & Butterfield, 1971; additional experiments are reviewed in Butterfield & Belmont, 1977, p. 283ff). Most pertinent to the present research, the usual conclusion is "strong evidence that subjects in self-paced procedures study difficult pairs longer than easy ones [defined in terms of the likelihood of being correct during test trials, not in terms of EOL]" (Zacks, 1969, p. 446).

A Model

Figure 1 shows a model that helps bring together the aforementioned components of (1) objective item difficulty, (2) EOL, (3) the allocation of study time, and (4) subsequent
memory performance. This model is not intended as a complete path-analysis diagram of psychological processing but rather as an organizational device within which to conceptualize past and present research.

The notion of objective item difficulty having a direct effect on a person's allocation of study time (cf. the previous quotation from Zacks, 1969; also echoed in Le Ny, Denhiere, & Le Taillanter, 1972) is indicated by the dashed arrow in the model; this direct connection might be an unconscious effect or, in any case, is characterized by a lack of any intervening influence from the person's belief about item difficulty. Another possibility, however, is that the relationship between objective item difficulty and the allocation of study time is at least to some degree determined by a two-stage mediational mechanism, whereby (a) objective item difficulty affects the person's belief about item difficulty (in accord with the aforementioned finding by Underwood, 1966) and (b) this belief, in turn, affects the allocation of study time. Put another way, objective item difficulty is monitored, and this component of the metacognitive system then is translated into the control of the allocation of study time. For instance, one simple translation rule would be that items with lower EOL will receive more study time, that is, an inverse relationship between EOL and the allocation of study time.

Notice that if extra study time is allocated to items believed to be harder, the overall result will tend to be a leveling in which the a priori harder items end up having recall that is raised in relation to that of the a priori easier items. In the limit, the correlation between EOL and subsequent recall might be zero if the person allocates study time so as to compensate completely for any differences in a priori item difficulty.

The model also includes motivation as a factor affecting the allocation of self-paced study time. In Figure 1, motivational instructions designed to raise the person's overall level of recall will presumably yield an increase in the total amount of study time that the person allocates, with or without affecting the amount of study time allocated to one item in relation to that of another item.

In accord with this presumption, we introduced the manipulation of speed-emphasized instructions (which emphasized that the person should study for only the minimum amount of time needed to learn the items) versus accuracy-emphasized instructions (which emphasized that the person should be sure to attain complete mastery of all the items, with no emphasis on studying quickly). Our initial reason for this manipulation was merely to assess the generality of conclusions about the relationship between EOL and the allocation of self-paced study time, that is, to establish whether the degree of relationship varied across small versus large amounts of overall study time. We believed we knew how those instructions would affect both the overall amount of study time and overall subsequent memory performance, namely, the accuracy-emphasized instructions (in relation to the speed-emphasized instructions) were expected to yield substantially greater overall study time that, in turn, would yield substantially greater overall memory performance (cf. Figure 1).

Hypotheses Investigated in the Present Research

Thus the focus of the present research, in terms of the empirical relationships we examined, is indicated by the lower half of Figure 1. The particular hypotheses under investigation, derived from the above sections, were the following:
Monitoring-affects-control hypothesis: Across items, there is an inverse relationship between monitoring judgments (e.g., EOL, FOK) and the amount of self-paced study time allocated to various items, that is, items lower in EOL or FOK will receive more self-paced study time.

Complete-compensation hypothesis: Across items, the relationship between the amount of self-paced study time and subsequent recall will be nil. This would occur if the extra study time allocated to one item, in relation to that of another item, compensates completely for any difference in difficulty between the two items. Such complete compensation has been reported for self-paced study time and subsequent recall of history versus science texts, with the participants allocating extra study time to the science texts and having equivalent subsequent recall for the two kinds of texts (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975).

Monitoring-neutralization hypothesis: Across items, the relationship between monitoring judgments (e.g., EOL, FOK) and subsequent recall will be nil when a person's study time is self-paced. This would occur if the relationships in the monitoring-affects-control hypothesis and in the complete-compensation hypothesis fully neutralize the perceived differences in item difficulty that would otherwise have been evident in recall (e.g., recall after experimenter-paced study trials; Underwood, 1966).

Instructions-affect-control hypothesis: The overall amount of self-paced study time will be substantially greater after accuracy-emphasized instructions are given than it will be after speed-emphasized instructions are given.

Instructions-affect-recall hypothesis: Overall recall will be substantially greater after accuracy-emphasized instructions are given than it will be after speed-emphasized instructions are given. We believed that this would result from the substantially greater self-paced study time after accuracy-emphasized instructions were given (ala the instructions-affect-control hypothesis), given the assumption that substantially greater study time would yield substantially greater recall. This assumption, which seemed presumable at the outset of the research, was disconfirmed, and we refer to the unexpected outcome as the labor-in-vain effect (discussed later).

Because our research yielded unexpected outcomes, especially regarding the instructions-affect-recall hypothesis, the generality and replicability of the conclusions seemed important. Therefore, we conducted three experiments, beginning with an investigation of nonsense-syllable items (so as to connect with the literature on EOL and experimenter-paced study time) and advancing to the more naturalistic items that are popular in current research, namely, general-information items that college students might try to learn in their day-to-day lives.

Experiment 1: Free Recall

Method

Subjects and Design

The subjects were 108 University of Washington undergraduates who volunteered for extra credit toward their psychology course grade. The design consisted of two groups: Speed of learning was emphasized for one group, and accuracy of learning was emphasized for the other group (verbatim instructions are reported under Procedure). Subjects were assigned randomly to the two equal-sized groups, according to order of appearance at the laboratory.

Items

The items were the same 27 trigrams used in the EOL experiment by Underwood (1966). These items varied greatly in inherent difficulty (e.g., BUG, FXH).

Apparatus

Items were presented via an Apple II microcomputer and were displayed on a cathode ray tube (CRT). The subject entered his or her response on the computer keyboard. Instructions, data collection, and statistical analyses were computer controlled.

Procedure

The experiment consisted of three discrete stages. First, the subject made EOL judgments on the 27 trigrams (EOL stage). Second, the subject studied the items during a self-paced study trial (study stage). Third, the subject recalled the items (recall stage). The procedural details for each stage are discussed separately.

EOL stage: The subject rank-ordered the 27 items in one of two ways, either from easiest to hardest or from hardest to easiest, to indicate how easy (hard) he or she believed the various items would be to learn. Items were presented simultaneously in a circular array on the CRT. The subject was instructed to examine the array and to indicate the item that he or she thought would be easiest (hardest) to learn. That item then disappeared, and the subject was instructed to indicate which of the remaining 26 items would be easiest (hardest) to learn. This procedure continued until all 27 items had been ranked. Subjects were given as much time as they needed to complete this process. The easy-to-hard versus hard-to-easy judgments were counterbalanced within each instructional group, such that half of the subjects in each group made the former and half made the latter.

Study stage: During study, the 27 items were presented in a random order. Each item remained on the CRT until the subject pressed the return key to indicate that he or she was ready for the next item. The subjects were told in advance that there would be only one opportunity to study each item. Half the subjects were given instructions that emphasized the speed of learning (speed group), and half were given instructions that emphasized the accuracy of learning (accuracy group). In the speed group, all the subjects were informed that we were interested in how quickly they could learn the items and that they were not to spend more time than they needed to learn each item. Each subject was told, "When we score your data, we will give you points for each of the trigrams that you recall correctly, but we will subtract points for each second that you spend studying the answer. The best way to score well in this experiment is to spend only as much time as is necessary to learn each trigram." In the accuracy group, every subject was informed that we wanted to determine only whether he or she could learn all of the trigrams. Each subject was told, "Do not press the return key until you have mastered the given trigram. When we score your data we will give you points for each of the trigrams that you recall correctly later in the experiment"; speed of learning was not even mentioned. The subjects were instructed to rehearse each item silently while it was on the screen; after study of the final item, 30 s of addition problems were presented to minimize recall from short-term memory.

Recall stage: The subject received a free-recall test for the 27 previously studied trigrams, and correct recall was defined as recalling the three letters of a given trigram in correct order (as in Underwood, 1966). Each subject was instructed to type as many trigrams as he or
she could remember in the 27 blanks. When the subject could no longer remember any trigrams, he or she was instructed to type the word *done*. The subject was then instructed to try for another minute to remember more trigrams. The recall phase ended when the subject typed the word *done* a second time.

**Practice trial.** After the subject was given instructions explaining the study stage, the arithmetic task, and the recall stage, he or she had a practice trial consisting of two items not used in the main experiment. The subject studied the items, computed sums for 30 s, and then attempted to recall the two items. After the practice trial ended, the subject received feedback concerning the time spent studying each item and the correctness of recall for each item. If the subject had no questions about the procedure, the main task began immediately.

**Results and Discussion**

Findings reported as significant had *p* < .05 unless otherwise indicated. All correlation coefficients are Goodman–Kruskal gamma correlations, which are the best available measures for this kind of data (for reasons, see Nelson, 1984; for interpretations and computational algorithms, see Nelson, 1986).

**Ease-of-Learning Judgments and Study Time**

The mean of the individual subjects’ correlations between EOL judgments and study time was −.32 for the speed group and −.27 for the accuracy group. Each of these correlations is reliably different from zero, confirming the monitoring-affects-control hypothesis that people allocate more study time to items that they believe will be difficult to learn. Moreover, this conclusion does not require an emphasis on either accuracy or speed of learning, and the small difference between groups was not reliable, *t*(106) = 1.53, *p* > .10. Notice, however, that these correlations are far from unity, which suggests that EOL judgments may not be the only factor affecting the allocation of self-paced study time.

**Study Time and Recall**

Contrary to the complete-compensation hypothesis, although the subjects allocated more self-paced study time to difficult items (i.e., items that subsequently they are still unable to recall), they did not allocate enough extra study time to eliminate the extra difficulty of those items. The mean correlation between study time and recall was −.25 for the speed group and −.20 for the accuracy group (i.e., items allocated more self-paced study time were still less likely to be recalled than were items allocated less self-paced study time). Each of these correlations is reliably different from zero, and they did not differ across the two groups, *t*(106) = .87. Although people allocate more study time to items they believe to be difficult (see previous paragraph), they apparently underestimate the amount of extra study time necessary for those items. On the one hand, this might suggest that people should allocate even more study time to difficult items. On the other hand, however, it may be more appropriate to ask why people bother as much as they do about allocating extra study time when they get so little extra recall in return (cf. the labor-in-vain effect discussed below).

**Ease-of-Learning Judgments and Recall**

A null or low correlation between EOL judgments and recall would confirm the monitoring-neutralization hypothesis. Instead, however, the EOL judgments were reliably correlated with subsequent recall accuracy, even though study was self-paced and the subjects could spend as much extra time as they wanted on items perceived to be more difficult: The mean correlation between EOL judgments and recall was +.48 for the speed group and +.48 for the accuracy group. Thus the subjects did not allocate study time in a way that completely compensated for perceived item difficulty. Moreover, the value of these correlations is identical to the one reported by Underwood (1966) between EOL judgments and rate of learning for these same items during experiment-paced study trials. A lower correlation might have been expected in the present experiment, because subjects were allowed to compensate (i.e., allowed extra study time) for more difficult items. However, this finding is consistent with the labor-in-vain effect (discussed in the next section); the extra study time that subjects allocate (via the monitoring-affects-control hypothesis) does not appear to yield substantial returns in terms of subsequent recall.

**Labor-in-Vain Effect**

As is indicated in the first row of Table 1, the median of the individual subjects’ median study time per item was reliably greater for the accuracy group than for the speed group, *U*(106) = 2417, *z* = 5.89. This shows that the instructional manipulation of speed of learning versus accuracy of learning was effective in influencing the amount of time that subjects spend studying the items (confirmation of the instructions-affect-control hypothesis).

The first row of Table 1 also shows that even though subjects allocated more than twice as much self-paced study time during acquisition under accuracy instructions as they did under speed instructions, the mean percentage of correct

| Experiment | Study time per item* | Percent correct recall*
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed emphasis</td>
<td>Accuracy emphasis</td>
<td>Speed emphasis</td>
</tr>
<tr>
<td>1. Trigrams</td>
<td>1.9</td>
<td>5.4</td>
</tr>
<tr>
<td>2. Word–trigram pairs</td>
<td>4.8</td>
<td>15.2</td>
</tr>
<tr>
<td>3. General information</td>
<td>1.2</td>
<td>8.4</td>
</tr>
</tbody>
</table>

* Median of individual subjects’ median study time (in seconds) per item.
* Mean of individual subjects’ percent correct recall.
* Reported data are for Trial 2; Trial 1 was experimenter-paced (5 s per item) and yielded 41% and 38% correct recall for the speed and accuracy groups, respectively (see text).
recall was quite similar for the two groups. The small difference (6%) was marginal and not even statistically reliable, t(106) = 1.92, .05 < p < .10, SEM = 2.41 and 2.30 for the speed and accuracy groups, respectively. Thus the subjects in the accuracy group gained little extra recall from their substantial extra study time.

We refer to this finding of substantial extra study time yielding little or no gain in recall as the labor-in-vain effect. Although we had anticipated that extra study time might yield diminishing (i.e., negatively accelerated) gains in recall, the present findings are quite extreme in showing not even a reliable gain in recall after more than twice as much extra study time. We might ask why the accuracy subjects even bothered to allocate this extra study time if they were gaining so little from it. The answer cannot be based on demand characteristics, because the instructions were manipulated between subjects, that is, the accuracy subjects did not know how much time the speed group spent studying, nor did they know that the speed group even existed! Moreover, in contrast to the situation of experimenter-paced study trials in which the subject may not try to learn during all of the available study time (which might be expected to produce little gain in recall from extra experimenter-paced study time), in the present situation the accuracy subjects presumably did try to learn during their extra study time. That is, if they did not want to spend extra time learning, they did not have to; they had control over the study time and could proceed immediately to the next item.

The first row of Table 1 also shows that recall was far from perfect for the subjects in the accuracy group. Even though those subjects had unlimited study time and were instructed to focus only on learning every item, they failed to recall about half of the items. These people did not accomplish the goal of perfect mastery, and the extra study time they did spend (in relation to the time spent by the speed-instructed group) was a labor in vain that yielded little or no extra recall.

**Experiment 2: Paired Associates**

Given some of the unexpected outcomes in Experiment 1, we wanted to investigate their replicability in situations other than free recall. Therefore, a paired-associate task was used in Experiment 2. This task also allowed for the collection of FOK judgments in addition to EOL judgments, thereby yielding a comparison of EOL and FOK as predictors of the allocation of self-paced study time. To maintain a connection with Experiment 1, we used the same trigrams as responses.

**Method**

**Subjects and Design**

The subjects were 72 undergraduates from the same population as in Experiment 1; none of them had participated in Experiment 1. The design consisted of two equal-sized groups that differed only in terms of the emphasis on speed of learning versus accuracy of learning. Subjects were assigned randomly to groups.

**Results and Discussion**

**Metamemory Judgments and Self-Paced Study Time**

The degree of relationship between self-paced study time and each kind of metamemory judgment (EOL and FOK) is shown in Figure 2 separately for each instruction group. Each correlation is reliably different from zero, replicating and extending the findings from Experiment 1. All are inverse relationships, showing that the items lower in either EOL or FOK tend to receive more self-paced study time, in accord with the monitoring-affects-control hypothesis.

**Relationship between the allocation of self-paced study time and EOL versus FOK.** One might expect that the allocation of self-paced study time would be related more to FOK judgments than to EOL judgments, because FOK judgments were made after some learning had already occurred (i.e., after the experimenter-paced study trials) and therefore would
reflect a more current state of knowledge. However, as Figure 2 shows, the allocation of study time is related at least as much to EOL judgments as to FOK judgments (for the speed group, $t(35) = 1.92, .05 < p < .10$; for the accuracy group, $t(35) = 1.44, p > .10$). This may have occurred because EOL judgments were made on the entire word–trigram pair, whereas FOK judgments were made with only the cue word available; subsequently during study, when the entire word–trigram pair again becomes available, subjects' reactions to the items may become more like their original EOL judgments.

Effect of speed emphasis in comparison with accuracy emphasis on the relationship between metamemory judgments and the allocation of self-paced study time. Figure 2 also shows a nonsignificant tendency, in the same direction as in Experiment 1, for EOL judgments to predict self-paced study time better for the speed group than for the accuracy group, $t(70) = 1.68, .05 < p < .10$. The relationship between FOK judgments and self-paced study time did not differ for the two instruction groups, $t(70) = .58$.

### Study Time and Recall

The correlation between self-paced study time and subsequent recall was $-.25$ for the speed group and $-.06$ for the accuracy group. The former correlation is reliably different from zero, in contrast to the complete-compensation hypothesis and as in Experiment 1 (see discussion there). The latter correlation is not reliably different from zero, in accord with the complete-compensation hypothesis. However, the difference between the magnitudes of these correlations is only marginally significant, $t(61) = 1.69, .05 < p < .10$.

### Metamemory Judgments and Recall

Figure 3 shows the degree of relationship between the two kinds of metamemory judgments (EOL and FOK) and recall performance. Various comparisons are possible by using the standard errors shown in the figure, but the focus here will be restricted to the issues discussed in the next three paragraphs.

**EOL and Trial 1 recall.** Figure 3 shows that EOL judgments reliably predicted recall after the experimenter-paced study phase for both the speed group and the accuracy group. These correlations are similar in magnitude to each other and to the $+.48$ correlation between individual subjects' EOL judgments and the rate of learning during experimenter-paced study trials in Underwood (1966).

**EOL and Trial 2 recall.** Although the correlation between EOL judgments and recall might have been expected to be nil after the self-paced study phase of Trial 2 (cf. the monitoring-neutralization hypothesis), Figure 2 shows that EOL judgments were reliably related to Trial 2 recall for both the speed and accuracy groups. This outcome replicates the corresponding findings from Experiment 1 and again disconfirms the monitoring-neutralization hypothesis. Also, the degree of relationship between EOL judgments and Trial 2 recall was reliably greater for the speed group than for the accuracy group, $t(61) = 2.00$.

**FOK and Trial 2 recall.** Figure 3 shows that FOK judgments were reliably related to Trial 2 recall for the speed group; however, for the accuracy group, the correlation was not reliably different from zero. To the extent that subjects allocated extra study time to lesser-known items, one might expect the relationship between FOK and subsequent recall to become neutralized (the monitoring-neutralization hypothesis). This is perhaps what occurred in the accuracy group, in which subjects had longer study times.

### Labor-in-Vain Effect

The second row of Table 1 shows the amount of study time and subsequent recall on each trial. Of primary importance, during self-paced study (i.e., Trial 2) the accuracy group studied the items more than three times longer than did the speed group (the difference was highly significant, $U(70) = 1091.5$, $z = 4.99$), but this substantial amount of extra self-paced study time yielded little extra recall. The small difference (8%) shown in the second row of Table 1 was not even significant, $t(70) = 1.42, p > .10$, $SEM = 4.25$ and 4.63 for the speed and accuracy groups, respectively. Thus, as in Experiment 1, there was a labor-in-vain effect in terms of little gain in recall from a substantial increase in self-paced study time.

Also as in Experiment 1, even though the accuracy group was instructed to learn every item, subsequent recall performance on Trial 2 was far from perfect (43% wrong). In contrast to Experiment 1, the present findings cannot be due to not knowing the demands of the recall test; here each person had the experience of being tested for recall on Trial 1, prior to having to decide when to terminate study during Trial 2.
Experiment 3: General-Information Items

The major findings from Experiment 1 were replicated and extended in Experiment 2. We wanted to determine if those findings were limited to nonsense-syllable responses (e.g., perhaps our students' unfamiliarity with that kind of item was critical for those results). Therefore we conducted one more experiment, examining the allocation of self-paced study time and subsequent recall for general-information items that the person could not already recall at the start of the experiment. These items were similar to what a college student might learn in college classes and everyday life.

Method

Subjects and Design

The subjects were 132 undergraduates from the same population as in Experiments 1 and 2 but who did not participate in either of those experiments. The design consisted of two groups that differed only in terms of the emphasis on speed of learning versus accuracy of learning, as in Experiments 1 and 2. Subjects were assigned randomly to the equal-sized groups.

Items

The items were 240 general-information questions that came from the Nelson and Narens (1980) norms and that are contained in the computer program FACTRETRIEVAL2 (Wilkinson & Nelson, 1984). For example, one question was, "What is the capital of Chile?" To ensure that the items were somewhat heterogeneous in difficulty, they were sampled from the overall pool of questions such that an equal number of incorrectly answered questions came from each of the following three ranges of normative probability of correct recall (Nelson & Narens, 1980): .000 to .081, .089 to .252, and .256 to .970. These particular cutoff values for the three ranges were chosen so that all three ranges contained the same number of expected recall errors if all 240 questions were asked.

Procedure

The experiment consisted of four discrete stages. First, each subject was asked general-information questions until 15 questions were not answered correctly (recall stage). Second, the subject made FOK judgments on the 15 nonrecalled items (FOK stage). Third, the subject studied the correct answers to those questions (study stage). Fourth, the subject was retested for recall of those answers (final-test stage). The procedural details for each stage are discussed separately.
Recall stage. Each question was presented individually on a CRT, and the subject was asked to "search memory hard in an attempt to find the answer." Subjects were encouraged to guess. However, if they had no guess about the answer, they typed next to proceed to the next question. To help minimize any effects of spelling errors, the computer evaluated the correctness of the subject's response by examining only the first three letters of that response.

The questions were presented randomly from one range of difficulty until an incorrect response occurred. Then questions were presented randomly from one of the two remaining ranges of difficulty until a second incorrect response occurred. Finally, questions were presented randomly from the remaining range of difficulty until a third incorrect response occurred. This cycle was repeated five times to produce the 15 nonrecalled items.

FOK stage. Each subject made FOK judgments on the 15 incorrectly answered items. Three questions were displayed simultaneously, and the subjects indicated their FOK by selecting whichever of those three questions they believed they were most likely to recognize the answer to. Then that question disappeared, and the subjects made a FOK judgment for the remaining two questions by selecting whichever question they believed they were more likely to recognize the answer to. These FOK judgments occurred for every combination of three questions, with the restriction that each question occurred with every other question exactly once. For $K$ items, this requires $K(K-1)(K-2)/6$ triads of questions, for a total of 35 triads for the 15 items. This procedure yields a FOK rank order (with the possibility of tied ranks) in terms of how many times a given item is chosen over all other items; the item chosen most frequently is assigned the highest FOK rank. To minimize an inappropriate high FOK for items to which the subject had guessed incorrectly during the recall stage (i.e., commission errors), the subject was informed that all of the items in the FOK stage either had a response of next during the recall stage or had been answered incorrectly (Krinsky & Nelson, 1985).

Study stage. The study stage consisted of two alternating tasks. The primary task was to study the correct answers to the previously incorrect 15 items. Presentation order of the 15 items for study was random. The subject saw a question for 5 s. Then the question disappeared, and a warning signal (four beeps and the word ready) appeared for 2 s. Then the correct answer appeared, and the subject controlled the amount of time that he or she wanted to study the answer; when the subject pressed the return key, the answer disappeared.

The secondary task was an arithmetic task that was inserted after the study of each answer to reduce short-term-memory effects and to eliminate any refractory phase prior to the next question and answer (Nelson et al., 1984). Arithmetic problems occurred for 30 s, and the cumulative number of problems correctly solved was displayed at the top of the CRT.

Final-test stage. The final-test stage was identical to the recall stage described above, except that only the items that the subject had just studied were tested. The order of presentation for the 15 questions was random.

Practice trial. After the subject received instructions explaining the study phase but before the study phase began, two practice trials occurred on items not used earlier in the experiment. After the practice trials, the subjects received feedback: Each subject in the speed group was shown his or her response, the correct response, and the study time for each item, as in Experiments 1 and 2. Each subject in the accuracy group was shown his or her response and the correct response for each item; but so as not to emphasize study time for this group, subjects were not informed of the study time they had allocated to each item.

Results and Discussion

The mean percentage of questions correctly recalled during the initial recall stage was 38% for the speed group and 35% for the accuracy group, showing that the two groups were similar in the ability to answer general-information questions, $t(130) = .93, p > .10$.

FOK Judgments and Self-Paced Study Time

The mean correlation between FOK judgments and the amount of self-paced study time was $-15$ for the speed group and $-06$ for the accuracy group. Each of these is reliably different from zero (in accord with the monitoring-affects-control hypothesis and replicating Experiment 2), but both are small in magnitude. This small magnitude does not appear to be due to insufficient range across items in each person's allocation of study time: The median (across subjects) shortest and longest study times were .4 s and 3.4 s, respectively, for the speed group and 2.4 s and 20.7 s, respectively, for the accuracy group. Also, the absolute magnitude of this correlation is reliably greater for the speed group than for the accuracy group, $t(130) = 2.27$, in contrast with those of Experiments 1 and 2, in which the difference was not reliable but was in the same direction as here.

Study Time and Recall

The mean correlation between study time and recall during the final-test stage was $-.25$ for the speed group and $-.21$ for the accuracy group. Each of these is reliably different from zero (contrary to the complete-compensation hypothesis). The difference between them is not reliable, $t(101) = .54$. This replicates the findings from Experiment 1 (see the discussion there).

FOK Judgments and Recall

The correlation between FOK judgments and recall during the final-test stage was $+.24$ for the speed group and $+.31$ for the accuracy group. These correlations are reliably greater than zero, indicating that subjects did not allocate their study time so as to completely compensate for perceived differences in item difficulty (contrary to the monitoring-neutralization hypothesis). This result extends the related findings from Experiments 1 and 2 to the learning of general-information items. These correlations did not differ reliably across the two groups, $t(101) = .83$.

Labor-in-Vain Effect

The third row of Table 1 shows the amount of study time and subsequent recall. During self-paced study, the accuracy group studied the items more than seven times longer than did the speed group (the difference was highly significant, $U(130) = 4221$, $z = 9.3$). However, this substantial amount of extra self-paced study yielded recall that was reliably but only slightly improved, $t(128) = 2.42$, $SEM = 2.17$ and 2.50 for the speed and accuracy groups, respectively. Thus, as in Experiments 1 and 2, there was a labor-in-vain effect in terms of little gain in recall from a substantial increase in self-paced study time.$^1$

$^1$ Previous research has shown that some FOK findings are different for commission-error items in comparison with omission-error items
Finally, as in Experiments 1 and 2, the subjects in the accuracy group had significantly less than perfect recall performance. They failed to recall 17% of the items, even though they were instructed to learn every item and were allowed unlimited study time to do so.

**General Discussion**

This discussion of the results is organized around the model shown in Figure 1 and the hypotheses listed in the introduction.

**Relationship Between Metamemory Judgments and Allocation of Self-Paced Study Time**

Every group of subjects in every experiment allocated their self-paced study time partly in accord with their EOL and FOK judgments. Items higher in EOL or FOK received less self-paced study time. This substantiates the two-stage mediation notion in the aforementioned model (see Figure 1) and confirms the monitoring-affects-control hypothesis. However, the magnitude of this relationship, especially with regard to FOK, is small.

One reason that FOK may be less predictive than EOL is that the former is stimulated by only the cue portion of the item, whereas the latter is stimulated by both the cue portion and the response portion (see Discussion section of Experiment 2). Meanwhile, the allocation of study time is an ongoing process and occurs while the person is seeing both the cue and response. Perhaps if FOK judgments were made while the person was seeing both the cue and response, then those judgments might be more predictive of the allocation of study time.

Another situation for which the degree of relationship might be greater is when the allocation of study time occurs not in terms of how much time to allocate before terminating study on a given item (as was examined here) but rather in terms of choosing which of several items to study (e.g., Masur, McIntyre, & Flavell, 1973). Rather than having our subjects choose one item over another item in advance, we used the procedure of having them decide how much study time to allocate to each item in an ongoing fashion, because that seemed more similar to what occurs in a naturalistic situation (e.g., when a student is studying for an exam).

Still another situation for which the degree of relationship might be greater is during multistrial self-paced learning. For instance, Zacks (1969) found high correlations (approximately +.70) between the amount of study time allocated to various items and a posteriori item difficulty. That large degree of relationship may have been due entirely to the multistrial situation, because the mean study duration on Trial \( n + 1 \) for wrong items in comparison with correct items on Trial \( n \) was 7.1 s in comparison with 3.6 s, respectively. (It is well-known that college students strongly choose to study only the previously nonrecalled items [Masur et al., 1973] and that they are quite accurate at telling which items they previously did versus did not recall [Vesonder & Voss, 1985].)

Another metacognitive monitoring component that might be relevant for predicting the allocation of study time is a dynamic judgment of learning (JOL). Perhaps the person compares the moment-to-moment JOL with his or her goal of intended threshold of mastery for the item (also called the "norm of study" by Le Ny et al., 1972), such that when the JOL reaches the intended threshold for that item, the person will move on to the next item regardless of how much time he or she initially expected to allocate. Thus there may be a dynamic update of how much study time the person believes is necessary for a given item. If so, this mechanism would have to be added to the model in Figure 1.

Finally, perhaps there is some degree of direct effect of item difficulty on the allocation of self-paced study time (cf. Figure 1) that the person is unaware of (i.e., an unconscious effect). We leave such possibilities as a topic for future research.

**Lack of Mastery During Unlimited Study Time**

In every experiment, two findings recurred that demonstrate lack of complete mastery even when the amount of study time is unlimited: First, the subjects terminated study before they had learned the items well enough to be 100% correct in subsequent recall. Second, they did not completely compensate for item difficulty by allocating enough extra study time to the more difficult items to learn those items as well as the easier items (disconfirming the complete-compensation hypothesis and the monitoring-neutralization hypothesis).

The finding of lack of mastery during self-paced study has occurred in other situations (Le Ny et al., 1972). Our explanation is this: While studying an item, if a person is able to recall the item from short-term memory, he or she may conclude that the item has been learned; however, subsequently, the person might be incorrect on that item during a test of long-term memory, even though earlier it could be recalled (from short-term memory). Put differently, the irony is that the person's metacognitive monitoring was valid (after all, he or she could recall the item during study, so in that sense it had been mastered), but that monitoring is partially irrelevant for the eventual task being studied for.

This could explain why metacognitive monitoring is sometimes misleading (e.g., yielding overconfidence in young children for the predicted amount of recall; Cavanaugh & Perl-
mutter, 1982) unless it is used in a more sophisticated way (e.g., perhaps by the person giving himself delayed tests that exceed the duration of short-term memory and that thereby can determine if the item is learned well enough to be recalled from long-term memory). A question for future research is, Does the person believe that the item is stored in long-term memory, or does he or she realize that the item is not stored in long-term memory but gives up because of not knowing what to do to get it stored there?

Incomplete compensation when allocating study time to difficult items is not always a poor study-time strategy. The strategy of having complete compensation, in the sense of equal likelihood of recall for every item, may not necessarily yield the best overall recall performance (e.g., in some situations it may be better to study only the easy items and sacrifice the hard ones, or at most devote only a little time to learning them). In our experiments, however, the person had unlimited study time to learn every item to complete mastery, and still the strategy adopted by our subjects did not yield complete compensation in the sense of level recall (i.e., equal likelihood of recall for every item, independent of the amount of study time allocated to each item).

Labor-in-Vain Effect

In all three experiments, the amount of self-paced study time was substantially longer for the accuracy-emphasis group than for the speed-emphasis group, ranging from twice as long in Experiment 1 to seven times as long in Experiment 3. However, this was labor in vain, because the two groups had little or no reliable difference in recall—no more than 8% in any of the three experiments. Put another way, the extra amount of self-paced study time in the accuracy group yielded only 1% or 2% more recall for each extra second of study time. Thus the relationship between study time and memory in the model in Figure 1 is not as potent as intuition might suggest, at least under the present conditions.

A related finding occurred in a free-recall experiment by Zimmerman (1975), in which the amount of self-paced study time was 13.6 s for two massed repetitions in comparison with 8.9 s for a single presentation, and the percentage of subsequent recall was 34% in comparison with 32%, respectively. Zimmermann remarked about the finding, "Indeed, subjects spent almost 50% more time studying a twice-presented massed item...than they did a single item. Yet, there was almost nothing to show for it in increased probability of recall" (p. 289).

A similar finding has also been common in the large literature on experimenter-paced study time in which extra study time has sometimes yielded little or no extra gain in subsequent recall. One example is the lack of a significant gain from massed repetitions in comparison with a single presentation, which is frequently explained by the attenuation-of-attention notion (i.e., the person does not bother attending to the second of two massed repetitions). Another example is extended experimenter-paced study time during a single presentation of each item; for example, Waugh (1970, Experiment 3) found no effect of the amount of experimenter-paced presentation time in a mixed-list design and explained her finding by supposing that the person studied a given item only as long as necessary to encode it and then ignored the item for the remaining presentation time (cf. the difference between nominal study time—the experimenter's presentation—and effective study time—the amount of time that the subject rehearses the item; Cooper & Pantie, 1967). Although the attenuation-of-attention explanation may be useful for experimenter-paced presentations, in which the subject must wait until the experimenter terminates the presentation of a given item, it does not explain the more striking finding of a labor-in-vain effect during self-paced study, in which the subject can terminate the presentation of a given item whenever he or she wants to.

What then is the explanation for the labor-in-vain effect during self-paced study? We do not know. What are the boundary conditions of the labor-in-vain effect? Again, we do not know, but we presume that there are some; after all, people learn entire lists! One possible boundary condition might be the kind of learning strategy that the person uses. For instance, Stoff and Eagle (1971) found that 1 s in comparison with 3 s of experimenter-paced study time had no effect on people who reported using a rote-repetition strategy (35% in comparison with 36% recall, respectively) but had a marked effect on people who reported using an organizational strategy (37% in comparison with 52% recall, respectively). Another possible boundary condition might be the stage of learning. For instance, would the labor-in-vain effect disappear during multilist learning or during the learning of multiple lists? Perhaps people need test trials and do not realize the ineffectiveness of their allocation/use of study time until after being tested (cf. Pressley, Levin, & Ghatala, 1984).

These questions for future research, and they are of practical as well as theoretical importance. The allocation of self-paced study time is what college students have to deal with in their everyday lives, and yet we know little about that control process. The present findings suggest that there is some interplay between the allocation of self-paced study time and the person’s metacognitive monitoring (EOL and FOK), but the remainder of the picture is far from complete.

References


