

Causes and constraints of the shift-to-easier-materials effect in the control of study

JOHN DUNLOSKY

University of North Carolina, Greensboro, North Carolina

and

KEITH W. THIEDE

University of Illinois, Chicago, Illinois

The shift-to-easier-materials (STEM) effect occurs when individuals adopt a low performance goal and subsequently select to restudy more easier items of a list than the difficult ones. The causes and constraints of the STEM effect were investigated across four experiments in which participants first briefly studied and judged their learning of 30 paired associates. They were then instructed to obtain a low performance goal and were asked to select items for restudy. The STEM effect was present when items were presented for selection simultaneously, but when items were presented for selection sequentially, the participants instead selected the more difficult items. The presence and absence of STEM effects were linked to planning that was triggered by the simultaneous format and to difficulties in executing an appropriate plan under the sequential format, respectively. Moreover, the STEM effect was evident for individuals with high memory spans but not for those with low memory spans. These and other findings highlight the contribution of planning and capacity constraints to the control of study time.

The issue of how individuals allocate study time across to-be-learned materials has been investigated increasingly in research on metacognition, stemming in part from the importance of this control process to student scholarship. To investigate the control of study time, researchers have typically used variations on a standard method. The procedure begins with a familiarity trial in which items—which may be paired associates, sentences, or texts—are presented individually at a fixed rate. At this time, participants are asked to rate different aspects of each item, including the ease of learning each item, the likelihood that each item will be remembered, and even one's interest in each item. Next, participants regulate study during a critical trial, either by pacing the study of each item and/or by selecting items for restudy. A common measure of metacognitive control involves correlating the judgments made during the initial trial with subsequent study time allocation, which partly describes the relationship between monitoring (as measured by the judgments) and control (see, e.g., Nelson & Leonesio, 1988). In a recent review of the literature, Son and Metcalfe (2000) noted that the modal—and nearly universal—outcome was an inverse correlation between people's judgments of learning (JOLs) and allocation of study time: Individuals are more likely to choose

for restudy (or to spend more time studying) items that are judged as least well learned than those judged as well learned.

Recently, however, Thiede and Dunlosky (1999) and Son and Metcalfe (2000) independently discovered circumstances in which individuals tend to allocate more study time to items judged as easiest to learn than to those judged as most difficult to learn. Consider results from Thiede and Dunlosky. During a familiarity trial, participants made a JOL for each of 30 paired associates. The items were then presented simultaneously in an array, and the participants selected those that they wanted to restudy. Prior to this selection phase, the participants had received one of two kinds of instruction. One group was instructed to do relatively well on the list, which had been the (explicit or implicit) instruction used in the majority of earlier investigations. As was expected, these participants selected most of the items for restudy and the correlation between initial JOLs and selection was negative, indicating that items selected for restudy were initially judged as relatively more difficult to learn. More important, participants in another group were told that they needed to learn only 6 of the 30 items. That is, they were given a low performance goal. These participants selected about 8 items for restudy, and the correlation was positive, indicating that they were more likely to select items for restudy that were judged as easier (vs. harder) to learn. This shift-to-easier-materials (STEM) effect contrasts the modal outcome, and, in doing so, represents one of the largest effects in the literature on the control of study. More important, it could not be readily accounted for by

We thank Chris Hertzog, Reed Hunt, Mike Kane, and Chris Lamb for discussions concerning this research, and Andy Conway for helping us obtain participants for Experiment 3. Correspondence concerning this article should be addressed to J. Dunlosky, P. O. Box 5190, Kent State University, Psychology Department, Kent, OH 44242 (e-mail: jdunlosk@kent.edu).

extant system theories of self-regulated study (for detailed discussion, see Metcalfe, 2002; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999).

In the study described above, the STEM effect was present when individuals selected items from an array in which all items were presented simultaneously for selection. By contrast, Thiede and Dunlosky (1999, Experiment 6) also had a group of participants select items for restudy as they were presented sequentially. Each item was presented individually, and the participants decided to study (or not to study) a given item prior to continuing to the next. Although the participants were instructed to obtain a low goal, they selected the majority of the items, and the correlation between JOLs and selection was negative. That is, these participants behaved as if they had been instructed to obtain a high goal. Why would the presentation format of items for selection—simultaneous or sequential—yield qualitatively different patterns of allocation? Answers to this question in particular, as well as explanations for the STEM effect in general, have not yet been empirically evaluated. Accordingly, a major goal of the present investigation was to explore some possible causes and constraints of the STEM effect.

An explanation of current interest involves the contribution of planning to the STEM effect. Planning may contribute to the effect under the simultaneous format in a rather straightforward fashion. That is, after receiving instructions to obtain a low goal, people may plan how to regulate study so that they would meet the goal of the task while minimizing the effort required to do so. Even though such a plan is quite simple, it could be a potent contributor to the STEM effect. To provide preliminary evidence about this planning hypothesis, we conducted a pilot study involving 51 college students who selected items using the simultaneous format. They first received a familiarity trial and made a JOL for each item. They were then instructed to obtain a low goal (i.e., learn 6 of the 30 items) and were presented with all 30 items simultaneously for item selection. Immediately after item selection, they were asked to respond in writing to the prompt, "Please describe in your own words how you decided to choose the particular items you did." We then scored a participant's answer as being indicative of planning to obtain the low goal if it endorsed choosing the easiest, most familiar items or if it explicitly mentioned choosing items to obtain the low goal. Thirty-eight of the 51 participants (75%) endorsed a plan that was in accord with effectively achieving the task goals. For each participant, we also computed a correlation between JOLs and item selection and the number of items selected. The mean correlation was larger in magnitude for those who endorsed planning (.30) than for those who did not (.04), although this difference was not significant [$t(40) = 1.01, p = .32$].¹ Reliably fewer items were selected by those who endorsed an appropriate plan (7.3) than by those who did not [12.3; $t(49) = 2.03, p < .05$]. These pilot data support the hypothesis that most participants

plan when they receive a low goal under a simultaneous format, and that planning influences item selection. Even so, because the reports were retrospective and may not be entirely veridical (Ericsson & Simon, 1980), we further evaluated the planning hypothesis in Experiment 1.

Why is the STEM effect absent when individuals select items for restudy under the sequential format? Across four experiments, we also empirically tested predictions that provide answers to this question. The predictions are based on two hypotheses, which are relevant to how people allocate study time after constructing a low performance goal for learning. Accordingly, item selection under a low performance goal will be our exclusive concern throughout this research, because it is here that individuals demonstrate qualitatively different patterns of regulation when items are presented using simultaneous versus sequential formats.² First, consider the lack-of-planning hypothesis. Under the sequential format with a low performance goal, planning may not occur, so an individual does not even consider choosing only a few of the easier items. Instead, individuals adopt the more typical allocation heuristic of focusing on relatively more difficult items (for exceptions to this typical heuristic, see Metcalfe, 2002). A second possible explanation of why the STEM effect is absent under the sequential format is provided by the inadequate-execution hypothesis. Even under the sequential format, individuals may plan to select the easier items for restudy, but fail to execute the plan appropriately. For instance, executing the plan to select a few of the easiest items for restudy may require cognitive resources to maintain the performance goal and plan, the relative difficulty of the items, the number of items already selected, and so forth. Under the sequential format, this kind of information must be kept active in working memory. Exceeding memory capacity may increase the likelihood that the plan will be poorly executed. Finally, both the lack of planning and ineffective execution of a plan may contribute to the absence of the STEM effect. That is, when items are presented under a sequential format, participants may not plan, but even if they do generate a plan, they may have difficulties executing it.

These and other hypotheses were investigated across the four experiments. Each one involved the same basic method: Participants were instructed to obtain a low performance goal and then to select items for restudy under either a simultaneous format or a sequential format. Prior to a description of each experiment, we discuss how the hypotheses were addressed.

EXPERIMENT 1

To isolate the source(s) of the difference in regulation under the two presentation formats, we experimentally evaluated the planning hypothesis—that planning contributes to the STEM effect—for selection under the simultaneous format. Two groups of college students participated in the standard procedure, which included a low performance goal. For both groups, the simultaneous

format was used to present items for selection. The instructed group was told to use the plan that was hypothesized as being used when no instructions were provided—that is, they were instructed to select six of the easiest-to-learn items for restudy. The control (or uninstructed) group was not given instructions to execute a plan. Our research strategy was to use the instructed group as a simulation of the uninstructed group. If uninstructed participants develop a plan to study the easiest items, the instructed group will simulate the behavior of the uninstructed group, resulting in instructional equivalence. Planned comparisons were conducted to evaluate whether both groups similarly selected few items and demonstrated the STEM effect.

Equally important, we included analogous groups that selected items under the sequential format. One group received the plan, and a control group received no instructions about planning. Evidence from these groups will also have implications for the aforementioned hypotheses. If the absence of the STEM effect results solely from not developing a rudimentary plan (i.e., to select six of the easiest items), then the instructed group is expected to demonstrate the STEM effect and choose few items. This outcome would be consistent with the lack-of-planning hypothesis. Of course, lack of planning and inadequate execution may both contribute to poor regulation. If so, the participants in the instructed group will have relatively intermediate values, yielding correlations that are smaller in magnitude (i.e., nearer zero) and selecting fewer items than will those in the control group. Finally, instructional equivalence may be obtained, with both groups demonstrating negative correlations and selecting many items, which would indicate that participants are unable to execute the plan successfully under the sequential format. These predictions were evaluated by planned comparisons involving the two sequential format groups.

Method

Participants and Design. One hundred six students from the University of Illinois at Chicago participated individually for class credit. The participants were assigned to one of four groups (simultaneous vs. sequential \times instructed vs. uninstructed) by order of appearance. Twenty-five students participated in each of the simultaneous format groups, whereas 28 students participated in each of the sequential format groups. The participants in the instructed groups were told to follow a plan for selecting items for restudy, whereas those in the uninstructed groups were not told about the plan.

Materials and Apparatus. The items consisted of 30 noun–noun paired associates. Fifteen were concrete–concrete pairs (e.g., *dog–spoon*), and 15 were abstract–abstract pairs (e.g., *democracy–gravity*). Both concrete–concrete and abstract–abstract pairs were used to increase the variability of difficulty across items. A Macintosh computer displayed all instructions and controlled all aspects of the procedure. The computers also recorded the participants' responses.

Instructions, List Construction, and Procedure. To increase the likelihood that the participants would attempt to attain the assigned goal, they were instructed that they would be playing a game that involved learning word pairs, the objective of which was to earn as many points as possible. The participants earned 10 points for

every item they correctly recalled on a test and earned 1,000 points if they exactly matched their assigned goal. Also, during restudy the participants lost 2 points for every second they spent studying an item. The participants in the instructed group were also told that when the goal was assigned they would be given a strategy to help them reach that goal; they were also instructed to use the strategy. During the instructions (which illustrated the game in detail), the participants were encouraged to ask questions about the procedure.

Items were randomized for study with the constraint that no more than two consecutive items could be concrete–concrete pairs or abstract–abstract pairs. During the familiarity trial, items were presented individually for 1 sec/item. Immediately after the final item of a list had been presented, the participants made a JOL for each item. JOLs were prompted with only the stimulus of an item (e.g., if *dog–spoon* had been presented for study, then the JOL cue would be “dog–?”) and the query, “How confident are you that in about ten minutes from now you will be able to recall the second word of the pair when prompted with the first? (0 = definitely won't recall, 20 = 20% sure, 40 . . . , 60 . . . , 80 . . . , and 100 = definitely will recall).”

After a JOL had been made for each item, the participants were shown their goal for the list, which was to correctly recall 6 of 30 items. For the participants in the instructed group, the instructions were as follows: “Your goal is to recall 6 of the 30 items. The best way to reach this goal is to study 6 of the items. Moreover, you should study the items that are easiest to learn.” For the participants in the uninstructed group, only the goal was presented.

Following assignment of the goal, the participants selected items for restudy. During selection, the participants receiving the simultaneous format were shown a 3×10 array (adapted from Masur, McIntyre, & Flavell, 1973) in which each cell was filled by a stimulus of a different item of the list. The cells were numbered from 1 to 30. The participants selected an item for restudy by typing the number of the corresponding cell. After an item had been selected, it was eliminated from the array. The participants typed the number 99 to indicate that they were done. The participants in the groups receiving the sequential format for item selection selected items for restudy individually. In particular, each stimulus of an item was presented individually with the query, “Do you want to restudy this item? Yes = 1, No = 0.” The participants in all four groups could select from 0 to 30 items.

For restudy, all of the items were randomized anew and re-presented. When a participant pressed the return key to begin a given trial, the first item was presented. The item remained on the screen until the participant again pressed the return key, which resulted in the presentation of the next item. We do not include any analyses of variables collected after item selection (e.g., self-paced study times and recall) because they were not relevant to our present aims and because self-paced study included all items instead of just those that were selected for restudy (as in Thiede & Dunlosky, 1999). Note, however, that the outcomes replicated those from our earlier research; more important, to foreshadow Experiment 4, we had the participants pace study for selected items only, which did allow an interpretable analysis of study time and study efficiency.

Presentation of the last item for restudy was followed by a 30-sec distractor task, which in turn was followed by paired-associate recall. For each recall trial, the stimulus of an item was presented, and the participants were instructed to type the correct response. Omissions were not allowed. After the final item was presented for recall, the number of points scored across the 30 items was presented.

Results and Discussion

For all four experiments, $p < .05$ for differences declared as reliable.

Relation between JOLs and item selection. For each participant, we computed a Goodman–Kruskal gamma correlation between JOLs and item selection. For item

selection, items that had been selected for restudy were scored with a 1, and items that had not been selected were scored with a 0. Thus, a positive correlation indicates that individuals more often selected items that were judged as easiest to recall. Means across individual gammas within each group are reported in Table 1. The planning hypothesis was evaluated by conducting planned comparisons between the two groups that selected items under the simultaneous format. The correlations from these groups were reliably greater than 0 ($ts > 3.0$) and did not differ reliably [$t(43) < 1.0$]. Thus, the STEM effect was evident for both groups.

To evaluate the hypotheses concerning why this effect does not occur under the sequential format, we conducted an analogous planned comparison, which revealed that the correlations under the sequential format were reliably different for the instructed and uninstructed groups [$t(47) = 2.85$], although neither group demonstrated a reliable STEM effect.

Number of items selected. We also conducted planned comparisons for the number of items selected for restudy. The mean numbers of items selected across individuals within each group are reported in Table 1. When selection occurred under the simultaneous format, the numbers of items selected for restudy did not differ reliably for the two groups [$t(48) < 1.1$]. Thus, both the correlational analyses and analyses of number of items selected demonstrated instructional equivalence, which is consistent with the assumption that planning contributes to regulation when participants are given a low performance goal with a simultaneous format. By contrast, when items were selected under a sequential format, the mean number of items selected for restudy was reliably less for the partic-

ipants who received the plan via instructions than for those who did not receive the plan [$t(54) = 2.84$].

Time prior to selection. The amount of time that elapsed prior to selection of each item provides descriptive information about the time course of selection. Although these analyses were exploratory, they provide further insight into how the individuals decided to select items for restudy. For instance, any plan development may require substantial time (e.g., for comparing the difficulty of items). If this is the case, more time will be used prior to item selection for groups that use a plan than for those that do not.

For each participant, we measured the time taken to select the first item for restudy and the time taken for selection between items in every consecutive pair thereafter. We then computed the mean time across individuals within each group (Table 2). We restricted our analyses to the first six selections because (1) the selection times reach an asymptote after about the fourth selection and (2) only about 60% of the participants selected more than six items for restudy, so that including more than six selections would have substantially reduced the number of observations in the analyses. We conducted a separate analysis of variance for those who selected under the simultaneous format and for those who selected under the sequential format.

When selection occurred under the simultaneous format, the main effect for group was not reliable [$F(1,40) < 1$]. The main effect for order [$F(5,200) = 903.6$] and the interaction [$F(5,200) = 2.4$] were reliable. Follow-up tests revealed that for both groups, time prior to selection was longer for the first item than for any of the subsequent items ($ts > 27.2$) and was also longer prior to selection of the second item than for that of any of the subsequent ones ($ts > 3.7$). Selection times for the remaining items did not differ reliably ($ts < 1.4$). Although speculative, our current interpretation is that this pattern reflects planning (prior to selection of the first item) and its subsequent execution, which itself occurs quite rapidly after a plan has been formed.

When selection occurred under the sequential format, decision times did not differ across groups [$F(1,54) < 1$] or across decisions [$F(5,270) < 1.5$], and the interaction was not reliable [$F(5,270) < 1$].

EXPERIMENT 2

The results from Experiment 1 provided converging evidence (with verbal protocols from the pilot study) for the planning hypothesis. The results also confirmed both the lack-of-planning hypothesis and the inadequate-execution hypothesis, which explain why the STEM effect did not occur when the participants selected items under the sequential format. In Experiment 2, we scrutinized these hypotheses in greater detail. First, consider the possible intricacies of planning for the present task. Although the instructed participants were given a plan, without being able to view the entire array of items a

Table 1
Relation Between Judgments of Learning (JOLs) and Item Selection, and Mean Numbers of Items Selected for Restudy

Experimental Groups	Gamma		Number Selected	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1				
Simultaneous format				
Plan instructions	+.44*	.14	8.6	1.3
No instructions	+.44*	.14	10.5	1.4
Sequential format				
Plan instructions	+.07	.14	14.4	1.8
No instructions	-.46*	.12	21.3	1.7
Experiment 2				
Simultaneous format	+.44*	.13	6.6	0.7
Sequential format	+.16	.16	11.2	1.9
Experiment 3 (Sequential Format)				
High-span participants	+.38*	.15	11.5	1.7
Low-span participants	-.12	.20	16.7	1.9
Experiment 4				
Simultaneous format	+.38*	.19	9.1	1.6
Sequential format	-.73*	.13	16.7	1.8

Note—Gamma, intraindividual gamma correlations between JOLs and item selection; a positive value indicates the STEM effect. Entries are means across individual scores. *The correlation is reliably different from 0 ($p < .05$, one tailed).

Table 2
Time Elapsed (in Seconds) Prior to Selection of an Item for Restudy

Experimental Groups	Order of Item Selection											
	1		2		3		4		5		6	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1												
Simultaneous format												
Plan instructions	30.9	1.3	8.2	0.4	6.1	0.4	6.1	0.4	6.4	0.5	6.1	0.5
No instructions	32.6	1.4	10.2	1.1	6.6	0.6	6.2	0.5	6.4	0.6	6.0	0.3
Sequential format												
Plan instructions	2.7	0.7	2.5	0.5	2.5	0.6	2.9	0.9	2.2	0.4	2.1	0.3
No instructions	2.5	0.3	2.2	0.2	2.1	0.2	2.1	0.3	2.1	0.3	2.1	0.3
Experiment 2												
Simultaneous format	39.0	8.9	13.2	1.5	10.3	1.6	9.8	2.3	8.8	1.1	8.3	1.8
Sequential format	7.7	0.9	5.0	0.8	3.0	0.8	2.2	0.3	1.8	0.2	2.5	0.4
Experiment 3 (Sequential Format)												
High-span participants	0.9	0.3	1.9	0.4	1.9	0.3	1.4	0.2	1.4	0.2	1.6	0.3
Low-span participants	1.5	0.6	3.0	0.5	2.6	0.5	2.7	0.7	2.5	0.6	2.7	0.7
Experiment 4												
Simultaneous format	29.0	3.4	10.1	3.4	7.4	2.1	5.3	1.3	4.3	0.8	4.1	0.7
Sequential format	5.7	0.5	3.5	0.5	2.8	0.5	2.9	0.5	3.0	0.5	3.1	0.4

Note—For order of item selection, 1 = first item selected for restudy, 2 = second item selected for restudy, and so forth. Data from 1 participant in Experiment 4 (simultaneous group) were excluded because the study time to the first selection was over 5 standard deviations above the group mean.

given participant may not consider which specific subset of items should be selected. That is, his or her plan may not be well specified, which in itself may undermine decisions about which items to select. By contrast, under a simultaneous format, participants may first compare items and covertly decide which subset to choose prior to overtly selecting any items. They will then proceed to select that subset of items from the array. This possibility is supported by the selection times of the group that received the simultaneous format. Second, the participants who received plan instructions under the sequential format may not have demonstrated the STEM effect because of difficulties in adequately executing the plan. That is, even if the participants had a relatively well specified plan, they may not have been able to execute it if doing so exceeded available cognitive resources.

To evaluate these hypotheses, we had one group of participants select items first under the simultaneous format and then under the sequential format. The rationale was straightforward: When participants select first under the simultaneous format, they form a plan and execute it adequately, which involves choosing a specific subset of items and subsequently selecting them for restudy. Because participants will first select the subset of items under the simultaneous format, if the absence of the STEM effect under the sequential format is due solely to inadequate planning, regulation under the sequential format will proceed effectively. However, if it does not occur because the participants have difficulty in executing the specific plan, then regulation under the sequential format will not be altered. That is, the STEM effect will not occur, and the participants will choose relatively many items for restudy. Note that this particular experiment favors the

former hypothesis, because it would seem rather trivial to initially select items under the simultaneous format and then select the same ones under the sequential format. Accordingly, an absence of the STEM effect under the sequential format would provide relatively strong evidence for the inadequate-execution hypothesis.

Method

Participants and Materials. Twenty-five students from the University of Illinois at Chicago participated individually for class credit. The items and other materials were the same as those used in Experiment 1.

Instructions, List Construction, and Procedure. All the participants completed the same procedure. The instructions and list construction were similar to those of the previous experiments (with no instructions about plans), except that the participants were instructed that they would be selecting items for restudy under a simultaneous format and a sequential format. The procedure was identical to that of Experiment 1, except that the participants selected items for restudy twice.

The participants were assigned a goal of correctly recalling 6 of 30 items and then selected items for restudy from the array. After they had finished selecting items from the simultaneous format, the goal was presented again, and they selected items for restudy under the sequential format. After selecting items for restudy a second time, the participants restudied the items and then completed a 30-sec distractor task and the test of paired-associate recall.

Results and Discussion

For each participant, we computed two Goodman-Kruskal gamma correlations between JOLs and item selection: one for the simultaneous format and another for the sequential format. Mean gammas were then computed across individuals for each group (Table 1). The gammas were reliably greater when the participants selected under the simultaneous format than when they selected under

the sequential format [$t(20) = 2.18$]. Moreover, a reliable STEM effect was evident only when the participants selected under the simultaneous format.

The mean number of items selected for restudy are presented in Table 1. Fewer items were selected under the simultaneous format than under the sequential format [$t(24) = 2.88$]. Thus, as in Experiment 1, when the participants had a low performance goal, they selected relatively few of the easier items for restudy. Perhaps surprisingly, even though the participants had just selected under the simultaneous format, the subsequent sequential presentation of items resulted in selection of more of the harder items for restudy. These results suggest that even when participants have a well-specified plan, the sequential format may undermine their ability to execute it successfully.

For each participant, we computed the mean decision time as in Experiment 1 (Table 2). Mean times were reliably greater for simultaneous selection than for sequential selection [$F(1,17) = 28.85$]. The time used prior to each selection also varied across decisions [$F(5,85) = 9.04$]. Furthermore, the interaction was reliable [$F(5,85) = 4.64$]. For both groups, the first selection took longer than the others; however, the difference was greater under the simultaneous format than under the sequential format.

EXPERIMENT 3

The results from Experiment 2 suggest that the sequential format may impede people's ability to adequately execute a relatively well-specified plan. As was alluded to above, limitations of working memory capacity may undermine decisions on how to allocate study. If so, when individuals are provided with the plan, its execution under a sequential format is expected to be more successful for those who have greater working memory capacity. In Experiment 3, we compared item selection between individuals with high memory spans versus those with low memory spans. Both groups were given a low performance goal and instructions to obtain this goal by choosing relatively few of the easier items for restudy (as in Experiment 1). Item selection then occurred under the sequential format. If memory capacity moderates successful plan execution, then in comparison with low-span participants, high-span participants would be expected to show relatively appropriate regulation, demonstrating the STEM effect and choosing fewer items for restudy.

Method

Participants and Materials. Forty students from the University of Illinois at Chicago participated individually for class credit. The items and other materials were the same as those used in the previous experiments.

Preexperiment screening of memory capacity and design. Prior to completing our experimental procedure, the participants had their working memory assessed via the operation-span task (Turner & Engle, 1989). This task involves remembering a series of

two to six words in order, while verifying the accuracy of mathematical equations such as $(8/2) + 2 = 6$. This span task involves a series of trials. During each trial, the number of equations presented may vary from two to six. After the final equation of a trial is presented, the participant must recall the words in their correct order. The number of equations changes randomly until the participant has completed three trials at each span. An individual's span is the sum of the number of correctly recalled words for trials in which recall of words was perfect and in the correct order. We used an extreme-groups design, with 18 participants scoring in the top quartile (operation span scores above 19) and 22 participants scoring in the bottom quartile (operation span scores below 11).

Instructions, List Construction, and Procedure. The instructions and list construction were similar to those of the previous experiments (with instructions about plans). All the participants completed the same procedure. After a familiarity trial, the participants made a JOL for each item. They were instructed that their goal was to recall 6 of 30 items and were given the plan for attaining the goal, as in Experiment 1. They then sequentially selected items for restudy, restudied the items, and completed a 30-sec distractor task and the test of paired-associate recall.

Results and Discussion

We computed a Goodman–Kruskal gamma correlation between each participant's JOLs and item selection. Mean gammas were then computed across individuals within each group (Table 1). The mean gamma was reliably greater for high-span than for low-span participants [$t(35) = 2.05$]. The mean number of items selected for restudy was reliably less for high-span than for low-span participants [$t(38) = 2.05$]. These results are the first to demonstrate that individual differences in memory capacity can influence the relatively fine-grained allocation of study time across items.

Given that selection occurred under the sequential format, we expected selection times to be relatively short for both groups. We also had no a priori expectations about the relation between span and selection times. Nevertheless, for completeness we report the means for decision times across individuals separately for each group (Table 2). The mean time increased slightly across selections [$F(5,190) = 4.59$]. Mean times prior to each selection did not differ between groups [$F(1,38) = 2.76$], and the interaction was not reliable [$F(5,190) < 1$].

EXPERIMENT 4

A motivating question for the present research has been that of why individuals demonstrate the STEM effect and choose relatively few items for restudy when they select items under a simultaneous format, but fail to do so under a sequential format. Inadequate planning and capacity constraints have been implicated by evidence described in the first three experiments. Even so, an alternative explanation for some of the findings is suggested by a self-handicapping hypothesis, according to which individuals choose to restudy the difficult items (resulting in the absence of the STEM effect) to ensure that they do not surpass the recall limit of six items. Although self-handicapping cannot readily account for

some results from the previous experiments (e.g., span effects in Experiment 3), the goal of correctly recalling exactly six items to receive bonus points may have provoked some of the participants to allocate study in ways that partly contributed to the absence of STEM effects under the sequential format.

We evaluated such possibilities in Experiment 4 by changing some of the instructions to make the procedure more closely reflect aspects of real-life goal attainment. Most important, in the previous experiments the participants received a point reward for obtaining the *exact* goal of six correct items, whereas students often strive to obtain *at least* a given grade. For instance, a student may study to earn a D or better on a final exam, so that she will be sure to obtain a B in class. Also, the particular reward used in the previous experiments was relatively extreme (1,000 points). In combination with the exact goal, the 1,000-point reward may have tempted a handful of participants to self-handicap during selection or to make up incorrect answers during the test. In Experiment 4, any participant who correctly recalled six or more items during the criterion test would obtain the reward, which was set at 60 points. The participants were not penalized for recalling more than six items. Thus, if self-handicapping was elicited in the previous experiment by the goal to correctly recall exactly six items, individuals who select under the sequential format in the present experiment will demonstrate the STEM effect and will select relatively few items for restudy.

Method

Design and Participants. Thirty students from the University of Illinois at Chicago participated individually for class credit. Fifteen participants were assigned to each of two groups (simultaneous format versus sequential format) by order of appearance.

Procedure. The procedure was identical to the one used in Experiment 1, except for three differences. First, the participants earned 60 points for correctly recalling six responses or more. The participants were instructed that if they learned more than the assigned goal demanded, they would be neither penalized nor rewarded; they would still receive the same number of points as if they had recalled exactly six responses. Second, the participants studied only those items that they had selected for restudy. Finally, the participants were instructed that they would lose 2 points for every second of self-paced study, and, as in the previous experiments, they were encouraged to earn as many points as possible.

Results and Discussion

Item selection. As in the previous experiments, we computed both the correlational measure of regulation (between JOLs and item selection) and the number of items selected for study, which are presented in Table 1. The correlation between JOLs and item selection was greater for the simultaneous than for the sequential group [$t(27) = 4.88$]. The number of items selected was less for the simultaneous than for the sequential group [$t(28) = 3.15$]. Also, as is shown in Table 2, decision times per item for the two groups correspond to those from Experiments 1 and 2.

Thus, under a new set of conditions, the key outcomes concerning item selection replicated (1) the STEM effect for the simultaneous group and the reverse, modal outcome for the sequential group and (2) the selection of fewer items by the simultaneous group than by the sequential group. Most important, these results establish that any self-handicapping elicited by instructing participants to obtain an exact goal of six items cannot account for the presence or absence of the STEM effect.

Total points, recall performance, and self-paced study. For completeness, we also present analyses of several measures that pertain to how many points the participants earned during the task. Sixty points were rewarded to a participant if he or she had correctly recalled six or more items, and 2 points were subtracted from the point total for each second of study. The number of points earned was greater for the simultaneous group ($M = 18.3$, $SEM = 6.6$) than for the sequential group [$M = -60.4$, $SEM = 34.0$; $t(28) = 2.27$].

In contrast with the previous experiments, self-paced study and recall performance were used to derive total points and, hence, analyses of these measures were also included. Concerning recall performance, the mean proportion correct did not differ for the simultaneous group ($M = .29$, $SEM = .04$) versus the sequential group [$M = .39$, $SEM = .06$; $t(28) = 1.30$]. The six-point goal was achieved by 12 of the 15 participants in each of the groups. For each participant, we computed the total amount of time used during self-paced study as well as the amount of time used per item studied. Although the mean study time per item did not differ reliably for the simultaneous ($M = 3.5$, $SEM = 0.7$) and sequential ($M = 2.2$, $SEM = 0.3$) groups [$t(28) = 1.64$], the mean total study time across participants was reliably shorter for the simultaneous group (18.8 sec; $SEM = 3.3$) than for the sequential group [57.2 sec; $SEM = 17.0$, $t(28) = 2.21$].

GENERAL DISCUSSION

With the present four experiments, we replicated key outcomes of previous research and demonstrated new phenomena that inform theory of self-regulated study. We now first discuss the relevance of the present outcomes to a hierarchical model of self-regulated study and then turn our attention to constraints on self regulation.

Hierarchical Model of Self-Regulated Study

To account for the extant data on self-paced study and item selection, Thiede and Dunlosky (1999) proposed a hierarchical model that consists of a "subordinate level in which the regulation of individual items occurs, along with a superordinate level that involves planning in which decisions about how to regulate study are made on the basis of how to complete the task as a whole" (p. 1025). The idea is that an individual first plans how to regulate study across all items of a list, and then control is trans-

ferred to a subordinate system in which individual items are studied. Although this hierarchical model provided a more adequate account of the extant data than did previous models that consisted of the subordinate (item-level) system alone, the assumptions of the model have not been previously tested.

The present research provides evidence consistent with the assumption that planning contributes to the STEM effect. When the participants had a low performance goal³ and selected items under the simultaneous format, instructions to select six of the easier items for restudy did not influence item selection. That is, focal behaviors of the group receiving the plan instructions simulated those of the uninstructed group. An alternative interpretation here is merely that the plan instructions were not efficacious; this alternative is inadequate, however, because providing the plan in another condition critically influenced selection (Experiment 1). Thus, the plan instructions were efficacious but did not influence behavior in a condition in which the plan was presumably being used.

With this interpretation in mind, however, the plan provided in Experiments 1 and 3 (i.e., to restudy about six of the easiest items) apparently underspecifies the one that is actually developed under the simultaneous format. In particular, consider the pattern of response times prior to selection of each item (Table 2). Individuals who were not given plan instructions did not select an item for restudy until about 30 sec had elapsed. When they were given the plan to restudy the six easiest items, a shorter preselection delay would have been obtained if the 30 sec were being used to develop this general plan. By contrast, results from Experiment 1 show that regardless of whether or not individuals are given plan instructions, the delays prior to the first selection are nearly identical. Our interpretation is that the majority of time spent prior to selection is used to develop a more detailed plan that involves covertly identifying the subset of items for subsequent selection.

The contribution of planning to self-regulated study also bears on another issue that has become central to theory of metacognition—namely, the degree to which metacognitive processes are implicit versus explicit (see, e.g., chapters in Reder, 1996). Reder and Schunn (1996) have argued that metacognitive monitoring and control processes often occur implicitly. For instance, in the case of self-regulated study, college students in our experiments would have many years of experience allocating study time differentially across items, so that allocation itself may be driven by an implicit heuristic. By demonstrating that planning can influence how individuals regulate learning of simple materials, the present research establishes that at least some metacognitive processes are explicit. This conclusion is also supported by a comparison of individuals with high versus low memory spans. In particular, Engle, Kane, and Tuholski (1999) proposed that “individual differences on measures of working memory capacity primarily reflect differences in capability for controlled processing and thus, will be

reflected only in situations that either encourage or demand controlled attention” (p. 104; see also Conway & Engle, 1994). Based on this rationale, the qualitative difference in selection between span groups suggests that the STEM effect partly relies on attentional control.

Constraints on Self-Regulated Study

In the introduction, we offered two hypotheses about why the presentation format qualitatively changed how individuals selected items when they had a low performance goal. Both were confirmed, with evidence implicating lack of planning (Experiment 1) and inadequate plan execution (Experiments 2 and 3) in the regulation of study under the sequential presentation format. We explore these issues briefly in the next sections.

Sequential format versus simultaneous format.

Previous effects of presentation format on self-regulated study motivated the present research by providing an impressive constraint on the control of item selection. Under the simultaneous format, individuals who have a low performance goal presumably plan and can effectively execute their plan. Under the sequential format, however, individuals fail to plan. These effects provide two avenues for future research, one of which is relevant to theory and the other to application.

Concerning the former, more theory-based avenue for research, why does presenting items simultaneously for item selection trigger planning? One possible explanation is that the simultaneous presentation of many items automatically engages comparison processes across items. Given that one must decide whether each item should be selected, comparisons across items may initially block progress toward making item-by-item decisions. The initial lack of progress may in turn prompt an individual to plan how to choose among the items. This sequence of processes is unlikely to occur for the sequential format because the action required to achieve the immediate goal of selecting an item is not blocked; one merely selects (or does not select) the item being presented.

With regard to application, the present findings provide clues on how to improve regulation. Planning apparently was disrupted both because the sequential format did not trigger it and because even when the participants had a plan it was challenging for them to execute it. Because successful execution of the plan presumably would require participants to maintain much information during selection, providing environmental support may improve the success of item selection. For instance, the current item could be displayed with other critical information, such as the judged difficulty of the item, how many items have already been selected, and so forth. By parametrically manipulating the presence of these and other factors, one could determine the kinds of information that are sufficient for supporting item selection. Finally, the general implications of the results for education are also intriguing and can be readily evaluated in future research. For instance, students may set relatively low per-

formance goals in a variety of circumstances, such as when they are under time pressure or when only a low score on a final exam will allow them to pass. In such contexts, regulation may be most effective if students first consider how to allocate study across all test materials simultaneously before they begin to study a subset of them.

Resource capacity. That resource capacity may influence the execution of plans was originally proposed by Miller, Galanter, and Pribram (1960):

[W]hen the decision is made to execute [the plan] . . . it is brought into the focus of attention, and as we begin to execute it we take on a number of menial but necessary tasks having to do with gathering data and remembering how far in the Plan we have progressed at any given instant, etc. . . . [W]e should like to speak of the memory we use for the execution of our Plans as a kind of quick-access, "working memory" (p. 65)

Evidence from Experiment 3 highlighted the contribution of working memory to plan execution by demonstrating that item selection was qualitatively different for high-span than for low-span individuals.

But why, specifically, were these differences in item selection obtained? One answer to this question originally motivated Experiment 3 and seems relatively non-controversial. Namely, in comparison with participants with high memory spans, low-span participants may be less skilled at actively maintaining all the critical information in memory. Another possibility is suggested by the view that working memory is a joint vestige of short-term memory plus controlled attention.⁴ For instance, Kane and Engle (2001) argue that high- and low-working memory-span individuals differ with respect to attention control processing "specifically in contexts that present potent competition between task goals and habitual responses" (p. 3). Perhaps individuals with low memory spans are able to maintain the critical information in short-term memory but did not do so because of deficits in controlled attention. The attention of such individuals may have been captured by the habitual response of merely selecting an item and, hence, they lost track of the overall plan to obtain the low performance goals under the sequential format. Adjudicating these alternative views promises to inform both theories of working memory and theories of self-regulated study.

Conclusion

To enhance achievement in everyday contexts in which people regulate learning, researchers will need to understand how individuals select materials for study. Evidence from the present research supported several new conclusions. First, the presence or absence of the STEM effect is in part caused by whether individuals engage in planning prior to item selection. Second, even if individuals have a plan, the ability to execute it may be undermined either by task demands or by limited memory capacity. Future research will continue to contribute to our understanding of self-regulated study by meeting

both theoretical challenges—such as understanding what triggers planning—and applied challenges—such as discovering how plans can be successfully executed even when resource capacity is taxed.

REFERENCES

- CONWAY, A. R. A., & ENGLE, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, **123**, 354-373.
- ENGLE, R. W., KANE, M. J., & TUHÓLSKI, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 102-134). Cambridge: Cambridge University Press.
- ERICSSON, K. A., & SIMON, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215-251.
- KANE, M. J., & ENGLE, R. W. (2001). *Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference*. Manuscript submitted for publication.
- MASUR, E. F., MCINTYRE, C. W., & FLAVELL, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrail free recall task. *Journal of Experimental Child Psychology*, **15**, 237-246.
- METCALFE, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, **131**, 349-363.
- MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. (1960). *Plans and the structure of behavior*. New York: Holt.
- NELSON, T. O., & LEONESIO, J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 676-686.
- REDER, L. M. (ED.) (1996). *Implicit memory and metacognition*. Mahwah, NJ: Erlbaum.
- REDER, L. M., & SCHUNN, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45-78). Mahwah, NJ: Erlbaum.
- SON, L. K., & METCALFE, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 204-221.
- THIEDE, K. W., & DUNLOSKY, J. (1999). Toward a general model of self-regulated study: An analysis of selection items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1024-1037.
- TURNER, M. L., & ENGLE, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, **28**, 127-154.

NOTES

1. One reason this trend was not statistically reliable was the fact that the majority of the participants' responses endorsed planning when they had to achieve a low goal under the simultaneous format (in accord with the planning hypothesis), which led to few observations for nonplanners. Perhaps ironically, the potential validity of the assumption would undermine the power of obtaining reliable results even with many participants. To sidestep this problem, we evaluated this assumption again in Experiment 1 using experimental methods.

2. The effects and conclusions described herein are not predicated on explicitly instructing students to obtain a specific low performance goal because other conditions that confront learners also will result in the construction of low goals. For instance, some students often delay studying for a test so long that they do not have enough time to master all the assigned material. With such time limitations, students are likely to adopt a low performance goal and, hence, allocate study to the more readily learned items. The important point is that many situations will provoke submastery goals, which may elicit planning. Even though the present questions could be answered using other methods, we chose to

instruct our participants to obtain a low goal because participants easily understand these explicit instructions. Accordingly, a failure to find the STEM effect would not be an artifact of participants' miscomprehending task goals.

3. One potential critique is that having participants adopt such a low goal of correctly recalling only six items would result in virtually no planning. Of course, to the degree that this observation is correct, the present outcomes would be even more compelling, because what presumably entails virtually no planning (vs. no planning whatsoever) has qualitatively different effects on people's allocation of study and apparently yields more efficient restudy.

4. Although the interpretation of working memory as (short-term memory plus) controlled attention does open avenues for future research, the validity of this interpretation is not critical to our general conclusion. That is, even if individual differences in capacity were attributable to encoding and retrieval skills, our main point is still viable: Individual differences in a central resource substantively influence how individuals select items for restudy.

(Manuscript received June 6, 2003;
revision accepted for publication December 1, 2003.)