Note-taking as a Strategy for Learning

Susan B. Trickett (stricket@osf1.gmu.edu)
George Mason University, Fairfax, VA 22030

J. Gregory Trafton (trafton@itd.nrl.navy.mil)
NRL, Code 5513, Washington, DC 20375

Abstract

We explore the effects of taking notes on problem-solving and learning in a scientific discovery domain. Participants solved a series of five scientific reasoning problems in a computer environment in which they had access to an on-line, unstructured notepad. The results show that participants who used the notepad performed better than those who did not use it. This improvement held even when these participants no longer used the notepad on subsequent tasks. However, not all uses of the notepad were equally effective; only those that involved deeper levels of processing were related to improved performance.

Introduction

At the heart of much scientific endeavor lies the scientific method—the systematic design of experiments to test hypotheses and the interpretation of the results of the experiments to assess the validity of those hypotheses. Mastery of the "scientific method" is considered crucial to the enterprise of science, because it applies across scientific domains.

However, many studies show that although some people conform to a normative model of scientific reasoning, many do not (e.g., Klahr & Dunbar, 1988; Trickett, Trafton, & Raymond, 1998). Instead, people frequently adopt other, less optimal strategies, such as conducting experiments without a hypothesis (Klahr & Dunbar) or even generating all possible experiments (Trickett et al.). It is a consistent result of such studies that people in general—children, college students, and adults alike—find scientific reasoning tasks hard and may fail to solve them altogether (e.g. Kuhn, 1989).

What can be done to support students, both as they engage in scientific reasoning tasks (performance) and as they learn to solve them without scaffolding (learning)? One might take a "systems" approach to bolstering performance. Several options come to mind: an intelligent tutoring system, partial scaffolding, a complex help system, to name a few. However, these are expensive, complex and time-consuming to build (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995). Another, less costly option is to focus on strategies rather than systems. For example, students who are taught strategies of self-explanation when studying problem examples have been shown to outperform those who do not (Chi, de Leeuw, Chiu, & LaVancer, 1994).

One general strategy that may help students learn is taking notes. Many studies have shown that students who take notes perform better than those who do not; however other studies have found no advantage for students who take notes (see Kiewra, 1985 for a review). Taken as a whole, the literature on note-taking shows mixed results.

In reviewing the findings of the note-taking literature, Kiewra (1985) suggests that the mixed results are due to the *kind* of note-taking participants engaged in. He argues that note-taking studies should focus on levels of processing during note-taking. Different note-taking strategies may vary considerably in the level of processing participants must engage in. Sometimes participants merely copy verbatim what is read or heard, involving only transcription (e.g., Laidlaw, Skok, & McLaughlin, 1993). At other times, participants engage in "conceptual note-taking" (e.g., Rickards & McCormick, 1988) or summarizing material (e.g., King, 1992), requiring some kind of filtering. Or participants may base their notes on some form of self-questioning, which involves some level of synthesis (e.g., Spires, 1993).

These different levels of engagement—"transcription," "filtering," and "synthesis"—can be understood in terms of theories of levels of processing within the psychological literature (Craik & Lockhart, 1972). Levels of processing research suggests that participants recall items better when they process material more elaboratively. Techniques that bring about deeper processing include generating elaborations (Bobrow & Bower, 1969), and using advance organizers and generating questions (Frase, 1975).

Viewed from the levels of processing perspective, we see that more elaborative note-taking strategies lead to better performance than more shallow strategies. The "transcription" level of note-taking corresponds to levels of processing which involve no elaboration. Indeed, when participants simply copy material from a text or lecture, note-taking does not result in better learning (Laidlaw, Skok, & McLaughlin, 1993). In fact, such note-taking is no more effective than underlining (Ayer & Milson, 1993). "Conceptual note-taking" and summarizing (the "filtering" level of note-taking) involve deeper levels of processing, and are more effective than merely copying material (Rickards & McCormick, 1988; King, 1992). Note-taking that involves self-questioning or reorganizing material (the "synthesis" level) maps directly to the elaborative self-questioning strategies implicated in superior performance on memory tasks. These note-taking strategies result in better performance than either copying verbatim (Spires, 1993; Shimmerlik & Nolan, 1976) or summarizing (King, 1992). In summary, the deeper the level of processing involved in the note-taking strategy, the greater and more stable the learning that results.

Typically, studies of note-taking have been conducted in traditional learning environments, such as classrooms and lectures. Little research has been done on the effects of note-taking on synthesizing information in problem-solving environments. But combining the note-taking and depth-of-processing approaches suggests that some kinds of note-taking might indeed be helpful in problem-solving, and that the strategy of taking notes—particularly certain kinds of notes—might help performance in ways other than simply improving memory for information. If so, providing a note-taking facility in computer-learning environments might be a relatively straightforward and inexpensive means of improving performance and learning on problem-solving tasks.

One computer-based problem-solving environment in which note-taking is supported is Smithtown, an economics microworld that is considered a scientific discovery learning environment (Shute & Glaser, 1990). Empirical studies of students using Smithtown have found that successful students made more notebook entries, overall, than less effective students (Shute & Glaser, 1990). However, Smithtown's notepad is highly structured. It contains both implicit and explicit instruction that not only prompts the student to take notes but also suggests how to set about doing so.

The type of note-taking supported by Smithtown does not appear to involve the deeper levels of processing discussed above. It is not known what effect, if any, more elaborative note-taking has on problem-solving performance. Perhaps more importantly for a learning system, it is not known how using such tools affects students' learning, that is, how well students perform on subsequent tasks when they might no longer have access to such a note-taking tool.

In this paper, we present a re-analysis of 3 studies in which students solved some simple scientific reasoning tasks in an environment which provided access to a note-taking facility that allowed them to take free-form notes. This re-analysis focuses on the note-taking behaviors of the participants and the relationship of this behavior to performance on the tasks. It provides important insights into the relationships among note-taking, performance, and learning.

Method

Three separate studies were conducted to investigate different issues in scientific reasoning. Two of the studies were carried out at the same time; the third was conducted the following semester of the same school year. In all 3 studies, participants performed the same tasks and used an identical interface. Because studies of students engaged in problem-solving tasks typically involve relatively small numbers of participants, and because there were only minor procedural differences among the studies, we combine data from all 3 studies in order to increase the power of our analyses.

Participants

Participants in all three studies were George Mason University undergraduates, who received course credit for their participation. There were 30 participants in each study—a total of 90 participants (42 males and 48 females).

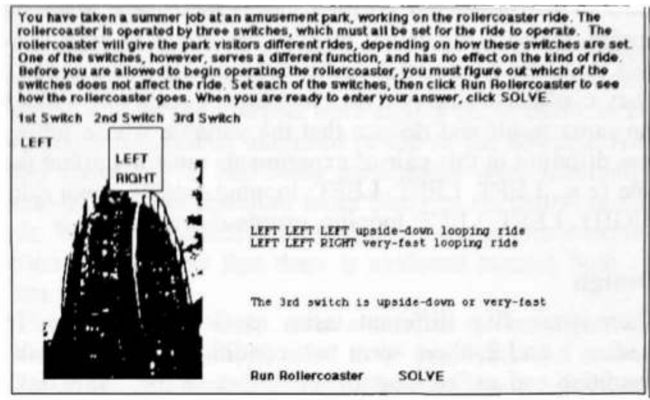

Figure 1: Screen snapshot of roller-coaster task

Materials

Five isomorphic scientific-reasoning tasks were developed, based on an adaptation of a task from Siegler and Atlas, (1976). For example, in one of the tasks, participants were told they were running a roller-coaster that was operated by three switches. The roller-coaster gave a different ride, depending on how the switches were set. All three switches had to be set for the roller-coaster to work; however, one switch did not affect the kind of ride. Participants had to identify the switch that did not affect the roller-coaster ride.

Each switch had two possible settings. Participants manipulated the setting of each switch and clicked on a "Run Roller-coaster" button to learn the kind of ride produced by that combination of settings. We refer to each new setting of the 3 variables, followed by clicking on the "Run Roller-coaster" button, as an *experiment*. Participants could run as many experiments as they wished before entering their solution. A record of each experiment and its results was displayed in a text box that remained visible throughout the task. If the text box became full, participants could use a scroll bar to view the results of their earlier experiments.

The interface also included a notepad, consisting of a blank text box, on which participants could enter information or comments if they chose. Figure 1 shows a screen snapshot of the interface for the roller-coaster task. The interface was the same for each task; only the instructions, variables, and answer were different across tasks.

Different cover stories were developed for four additional tasks. Instead of a roller-coaster, these tasks involved a musical instrument, a catapult, chemicals, or genetics (no domain knowledge was required for any of the tasks). The tasks were isomorphic in that they shared the same deep structure and could be solved by applying identical procedures (Simon & Hayes, 1976). For each task, there were three possible causal variables, each with two levels. One variable had no effect, and the goal was identify that variable.

Analysis of the problem space identified several different strategies by which the tasks could be solved. Participants could test each variable in turn by changing its setting while holding the other two constant—the optimal vary-one-thing-at-a-time strategy, or VOTAT (Tschirgi, 1980). They could identify the effect of each level of each variable (e.g., the third switch in the left position makes the roller-coaster go upside-down). They could identify the effect of each variable (e.g., the third switch makes it go upside-down or fast). They could also find two different experiments that yielded the same result and deduce that the variable whose setting was different in this pair of experiments must not affect the ride (e.g., LEFT, LEFT, LEFT: looping, upside-down ride; RIGHT, LEFT, LEFT: looping, upside-down ride).

Design

There were five different tasks, as described above. In studies 1 and 2, there were two conditions, a "same task" condition and an "isomoporh" condition. In the "same task" condition, participants were asked to solve the same task five times. In the "isomorph" condition, participants were asked to solve the series of five different, isomorphic tasks. In Study 3, all participants solved the series of five different tasks, i.e., were in the "isomorph" condition. In *both* conditions, the correct solution for a task was randomly generated for each task. In all three studies, the interface for the tasks was identical. All participants had access to the notepad, but were neither encouraged to use it nor discouraged from doing so.

Measures

Keystroke data, including entries participants made on the notepad, were collected as participants solved the tasks. In addition, in Study 1, verbal protocols were collected.

We used keystroke data to determine the accuracy of each participant's solution for each task. In order to investigate the use of the notepad, we identified three patterns of notepad use and coded each participant on each task as follows. Each task on which the participant made an entry on the notepad was coded as a *use* of the notepad. We were also interested in any possible carry-over effect of using the notepad. Consequently, after a participant used the notepad on one or more of the five tasks, each subsequent task on which they did *not* use it was coded as a *scaffolded non-use* of the notepad. All other tasks on which the participant did not use the notepad were coded as *non-use* of the notepad (i.e., tasks for which participants did not use the notepad or had not used it on a previous task). Table 1 illustrates this coding scheme.

Subj	Task	Use Notepad?	Notepad Code
100	1	No	Non-use
100	2	No	Non-use
100	3	Yes	Use
100	4	Yes	Use
100	5	No	Scaffolded non-use

Table 1: Codes for whether notepad was used

In addition to coding whether participants used the notepad, we coded *how* they used it. Several types of notepad entry were identified, as follows. Some participants identified what the variables did (IV). This category includes identifying the effect of each *level* of a variable and identifying the effect of the *whole* variable. Some participants noted that two different experiments yielded the same result (2-same). These two uses of the notepad map directly to the strategies outlined above by which participants could successfully solve the task. Some participants used the notepad to represent or re-represent the experiments they had run (RE). Some participants stated a hypothesis (SH). Entries on the notepad that engaged the task but did not fit any of these categories, that merely typed text that was visible on the screen, and/or were used by only one participant were coded as "Other" uses. Notepad entries that demonstrated more than one category of use were coded as "Mixed" entries.

There were nine entries on the notepad that did not fit any of the categories described above. These entries did not pertain to the task participants were asked to solve; for example, they were comments to the experimenter. Because there was no connection between these entries and the problem-solving task, these uses were discarded and recoded as *non-use* of the notepad. Table 2 summarizes the coding scheme for types of entry and gives examples of each category.

Type	Example
IV	Yerk-bubbly; Anjo-green; Ilop-hot Second switch affects looping or backwards
2-same	RRR & RRL same ride; LLL & LLR same ride
RE	1+1+1 = green glowing 1+2+2 = bubbly green
SH	It may be the second switch.
Other	Green-HIG BIG; Clear-BYA HYA
Mixed	Ivory piece is there or not Plastic mouthpiece makes it treble
Discard	If the answer is not chromosome 5 then these surveys should be deemed ineffective ...

Table 2: Codes for type of notepad entry

Finally, among participants who used the notepad, we identified the number of experiments they had conducted when they first used the notepad for a task. This number could range from 0 (if a participant used the notepad before beginning any experimentation) to the total number of experiments run (if they used it at the end of experimentation).

Procedure

Participants were trained on the interface. They practiced designing and running experiments, viewing the results, and using the notepad. They were told that they did not have to use the notepad, but should do so if they wished.

Results and Discussion

Although all 3 studies were experimental, each was designed to explore different issues in scientific reasoning. Thus the results presented in this paper are correlational in nature and are subject to the usual caveats in interpreting correlational data. However, we find evidence within the studies that supports a broader interpretation, as we discuss below.

We first analyzed the extent to which participants used the notepad in each of the three studies. In each study, there were 30 participants, who each performed 5 tasks, i.e., there were 150 tasks or opportunities for notepad use. In Study 1, there were 26 uses of the notepad (17% of tasks); in Study 2, there were 16 uses (11% of tasks); and in study 3, there were 40 uses (27% of tasks). In all 3 studies combined, there were 82 uses of the notepad over 450 tasks (18% of tasks).

In order to ascertain that there were no quantitative differences between conditions in terms of notepad use, we performed an ANOVA comparing use of the notepad by condition (same-task, isomorph). The results of this analysis were non-significant, F (1, 448) = 1.23, MSE = .54, p = .26, with means of .42 and .5, respectively. This suggests that participants in the same-task condition were neither more nor less likely to use the notepad than those in the isomorph condition. Therefore, because the number of notepad uses in each individual study was rather small, we combine the results across the three studies in all subsequent analyses.

Use vs. Non-Use of Notepad

In order to investigate whether using the notepad had an effect on problem-solving performance, we performed an ANOVA comparing participants' correct solutions in *non-use* of the notepad with correct solutions in *use* of the notepad. The result was significant, F (1, 382) = 3.99, MSE = .22, p < .05. (Means were for .7 *non-use* and .86 for *use*). This result suggests that participants who used the notepad, regardless of how they used it, were more likely to solve the problem correctly than participants who did not use the notepad. It appears that using the notepad was associated with better performance in this type of problem-solving.

Scaffolded Non-Use of Notepad

Next, in order to explore the relationship between use of the notepad and learning, we investigated whether there was an effect of the *scaffolded non-use* of the notepad. Recall that *scaffolded non-use* of the notepad refers to those tasks on which participants did not use the notepad but had used it on prior tasks; that is, it refers to tasks on which they *no longer* used the notepad. *Scaffolded non-use* of the notepad could not occur on the first task in the series of five tasks that participants performed and was more likely to occur later in the series. Because in general performance improved as participants proceeded through the series of tasks (Trickett, Trafton, & Raymond, 1998), we performed an ANCOVA comparing correct solutions in *non-use* with correct solutions in *scaffolded non-use* of the notepad, with task as covariate. The result of this ANCOVA was significant, F (1, 365) = 4.04, MSE = .21, p < .05. (Means were .65 for *non-use* and .82 for *scaffolded non-use*).

The result of this analysis suggests that participants who did not use the notepad but had used it earlier were more likely to solve the problem correctly than participants who never used the notepad. Thus, using the notepad can be seen to serve a scaffolding function, as the benefit of using the notepad carries over to later tasks even when a participant no longer uses it. This result shows an association between using the notepad and improved learning on these tasks.

Issues of Self-Selection

As discussed above, these data are correlational. One plausible interpretation of the results is therefore that participants who used the notepad were simply more likely to be successful for reasons unrelated to use of the notepad. Another possibility is that some participants were inherently "note-takers" and therefore more likely to do better on the task. While we cannot entirely reject these possible explanations, we believe that there is evidence against both of them.

First, recall that there were a number of strategies by which participants could solve these problems. As an explicit hypothesis-testing strategy, the varying-one-thing-at-a-time (VOTAT) strategy is considered the optimal strategy in scientific reasoning tasks (Vollmeyer, Burns, & Holyoak, 1996). If participants were "naturally good" at scientific reasoning tasks, we would expect them to use the VOTAT strategy. Contrary to this expectation, however, in Study 1, at least, there were very few instances of VOTAT (Trickett, Trafton, & Raymond, 1998)[1]. Instead, in that study, accuracy was highly correlated with systematicity.[2]

Second, given that systematicity was strongly related to accuracy on these tasks, we might expect participants using the notepad to be more systematic than those who did not. However, this was not the case; there was no significant correlation between systematicity and notepad use in Study 1. Thus we find two distinct strategies (systematicity and notepad use), both of which correlate with successful performance, but which do not correlate with each other. This result, coupled with the general lack of the VOTAT strategy, suggests that notepad users were not just better students.

Third, if some participants were simply more inclined to take notes, we would expect the majority of entries on the notepad to occur very early in the problem-solving process, that is, after participants had run only a very few experiments. However, again, this was not the case. Across all 5 tasks, only 16 (19.5%) of the 82 notepad entries were made in the initial stages of problem-solving (after 0 to 2 experiments). Participants used the notepad for the first time in the middle stages of problem-solving (after 3 to 5 experiments) on 30 tasks (36.5%), and in the late stages of problem-solving (after 6 or more experiments) on 36 tasks (44%). Table 3 shows a complete breakdown of these results.

# experiments	# first uses	% first uses
0	7	8.5%

[1] Use of the VOTAT strategy could only be determined in Study 1, because verbal protocols are needed in order to ascertain that this strategy is being used. We do not expect, however, that use of the VOTAT strategy would have been greater in studies 2 and 3; note, for example, that there were no uses of VOTAT recorded on the notepad in any of the studies, whereas all the other strategies were represented on the notepad.

[2] Systematicity refers to the entire set of experiments a participant generated for a task. Participants' data collection was coded as systematic if at least 75% of their experiments conformed to a discernible pattern.

1	5	6%
2	4	5%
3	13	16%
4	10	12%
5	7	8.5%
6	7	8.5%
7	5	6%
8	19	19%
9+	5	6%

Table 3: Experiments run at first use of notepad

These results show that by far the majority of participants did not approach these tasks already equipped with the strategy of using the notepad. Rather, they appear to have run several experiments and *then* turned to the notepad. Particularly interesting is the fact that the largest number of first uses occurred after participants had run 8 experiments. There were 8 possible different experiments that could be run. 19% of first notepad uses occurred when participants had run 8 experiments. It seems likely that, having exhausted all possibilities for collecting data, participants used the notepad to try to make sense of these data. Certainly, it does not appear that participants were *a priori* inclined to take notes.

How the Notepad Was Used

The analyses described above show that participants were more accurate if they used the notepad and that this advantage was maintained even if they subsequently stopped using it. However, the analyses do not differentiate among different types of notepad use. They do not address the question of whether some uses of the notepad were more effective than other uses. To investigate this question, we examined type of notepad entry in relation to the accuracy of the solution.

Because the number of notepad entries for some categories was quite small, it was not appropriate to perform statistical tests to determine the effectiveness of each of the different ways of using the notepad. Instead, we report the percentages of tasks with correct solutions for each type of entry. Table 4 summarizes the results of this analysis. The findings show that if participants did not use the notepad, they were correct 65% of the time (baseline). Four types of entry on the notepad (identify variable, note 2-same, re-represent experiments, and state hypothesis) were at least 18% above baseline. Two types of entry, "Other" and "Mixed," were below baseline.

Entry type	#correct	total	% correct
None	197	302	65%
Identify effect of variable	35	38	92%
Note 2 results the same	5	6	83%
Re-represent experiments	10	11	91%
State hypothesis	7	8	88%
Other	1	10	10%
Mixed	5	9	56%

Table 4: Percent correct per type of entry

Thus it would appear that use of the notepad *per se* is not linked to more successful performance. Instead, some uses of the notepad seem to be more helpful than others. Specifically, using the notepad to identify the effect of a variable, note two identical results, re-represent experiments, or state a hypothesis seem related to successful performance. We thus reclassified these four uses as "good" uses of the notepad. On the other hand, using the notepad for uses coded "Other" or mixing uses of the notepad seems to be associated with unsuccessful performance. We thus reclassified these two uses as "poor" uses of the notepad.

In order to test whether "good" uses of the notepad were more likely to lead to successful performance than "poor" uses, we conducted a Fisher's exact test. In doing so, we violate one assumption of this test, namely that the observations were independent (they were not independent, because each participant performed 5 tasks). However, such violations usually lead to a more conservative test. The results of the Fisher's exact test (two-tailed) were *significant*, $p < .001$.

Thus it seems that the four "good" uses are more likely to be associated with successful performance than the two "poor" uses of the notepad. Why might this be the case? We believe it is because the four good uses all involve a deeper level of processing. These helpful uses—identifying what a variable does, identifying 2 identical results, re-representing experiments and stating a hypothesis—all require reorganization or synthesis, both of which involve deeper processing than simply copying what is on the screen. This deeper level of processing would not be involved in the "Other" uses of the notepad, nor perhaps in the "Mixed" uses, depending on the combination of uses involved. Furthermore, identifying what a variable does and identifying 2 identical results correspond to strategies by which the tasks could be solved. Possibly, if participants were having difficulty solving the task, using the notepad might have been instrumental in helping them to develop a good strategy for the tasks.

Concerning the "poor" uses, several observations can be made. First, recall that the "Other" category included instances where the participant just re-typed some text that was visible on the screen. The low success-rate for participants using the notepad for "Other" uses (10%) suggests that using the notepad in ways that involve shallow processing did not help. Second, some "Other" uses of the notepad were idiosyncratic, in that they could not be categorized according to the general types of entry identified among the majority of participants. Thus it may be that these "Other" uses indicate confusion on the part of the participants. It seems plausible to think that they turned to the notepad in an effort to do something to move their problem-solving forward. However, because they had no clear idea about how to use the notepad effectively, it was of no benefit in these cases.

Finally, in 8 out of the 9 cases of "Mixed" entry, one of those uses was "Other." On 5 of those 8 tasks, the "Other" use came first and was followed by the helpful use. Participants were correct on 4 of those 5 tasks. If, the "Other" use is evidence of floundering, as suggested above, this shift in type of notepad entry might be an indication of a transition point in their problem-solving. Although there are too few instances to allow us to draw firm conclusions, the data and the trace of problem-solving activity provided by the notepad entries suggest that these participants were able to recover from their confusion and move to a good solution strategy.

General Discussion

The results described above show that participants who used the notepad were more likely in general to solve the problem correctly than those who did not use it. Moreover, this advantage was maintained even if participants stopped using the notepad on subsequent tasks. Although these data are correlational, there is some evidence to suggest that participants who used the notepad were neither "better" students nor for some reason intrinsically inclined to use the notepad. In addition, only uses of the notepad that involve deeper levels of processing were associated with successful performance; those that involve shallow levels were not.

These findings extend the current research on note-taking during problem-solving in a number of ways. First, the general finding that participants who used the notepad were more successful than those who did not suggests that taking notes can indeed be a helpful strategy in a problem-solving domain. It appears that the benefits of taking notes extend beyond boosting the simple recall of learned material, to helping students make sense of data they have generated and possibly to helping them to develop good problem-solving strategies.

Second, we suggest that the reason some uses of the notepad were related to better performance while others were not is that the "good" uses engaged the participants in deeper, more elaborative processing than the poor uses. We further specify those kinds of note-taking that involve deeper processing.

Third, the finding that the benefit of using the notepad carried over to tasks on which participants no longer used it suggests that taking notes can actually help students *learn*. This finding further supports our interpretation that using the notepad helped participants develop robust problem-solving strategies that they continued to apply across tasks. Having developed a sound strategy, participants most likely no longer needed the scaffolding provided by the notepad but could continue their problem-solving efforts independently.

Taken together, the results suggest that having students take notes that involve deeper processing as they solve scientific reasoning tasks is one means by which their performance might be improved. Providing a note-taking facility for students engaged in these types of problems is an inexpensive and comparatively straightforward form of scaffolding that may have the further advantage of helping students learn general strategies by which these problems can be solved.

As we have mentioned, these conclusions remain tentative, because of the correlational nature of the data. We cannot decisively reject other possible explanations, such as that the notepad users were "better" problem-solvers, although we find evidence to suggest that this is not the case. Clearly, our next step is to test these results experimentally, by manipulating use of the notepad and providing the appropriate controls. This experiment is currently being run.

Acknowledgments

This research was supported in part by a student fellowship from George Mason University to the first author and by grant number 55-7294-A8 from the Office of Naval Research to the Naval Research Laboratory. We wish to thank Erik M. Altmann, Robert W. Holt, Irvin R. Katz, Audrey W. Lipps, Sheryl L. Miller, Llelyn D. Saner, and Christian D. Schunn for their comments. We give special thanks to Paula D. Raymond for her continuing comments on this research.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167-207.

Ayer, W. W., & Milson, J. L. (1993). The effect of notetaking and underlining on achievement in middle school life science. *Journal of Instructional Psychology,*] *20*(2), 91-95.

Bobrow, S. A. , & Bower, G. H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology. 80*(3, Pt. 1), 455-461.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Frase, L.T. (1975). Prose processing. In G. H. Bower (Ed.), *The psychology of learning and motivation,* (Vol. 9). New York: Academic Press.

Kiewra, K. A. (1985). Investigating notetaking and review: A depth of processing alternative. *Educational Psychologist, 20*(1), 23-32.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational research Journal, 29*(2), 303-323.

Klahr, D., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science, 12,* 1-48.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689.

Laidlaw, E. N. Skok, R. L., & McLaughlin, T. F. The effects of notetaking and self-questioning on quiz performance. *Science & Education.. 77*(1), 75-82.

Rickards, J. P. , & McCormick, C. B. (1988). Effect of interspersed conceptual prequestions on note-taking in listening comprehension. *Journal of Educational Psychology, 80*(40), 592-594.

Shimmerlik, S. M., & Nolan, J. D. (1976). Reorganization and the recall of prose. *Journal of Educational Psychology, 68*(6), 779-786.

Shute, V. J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discover world: Smithtown. *Interactive Learning Environments, 1*, 51-77.

Siegler, R. S. & Atlas, M. (1976). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Detecting interactive patterns in data. *Journal of Educational Psychology, 68*(3), 360-370.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: problem isomorphs. *Cognitive Psychology, 8*, 165-190.

Spires, H. A. (1993). Learning from a lecture: Effects of comprehension monitoring. *Reading Research & Instruction, 32*(2(, 19-30.

Trickett, S. B., Trafton, J. G., & Raymond, P. D. (1998). Exploration in the experiment space: The relationship between systematicity and performance. *Proceedings of the 20th Annual Meeting of the Cognitive Science Society* (pp. 1067-1072). Mahwah, NJ: Erlbaum.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1-10.

Vollmeyer, R., Burns, B. D. & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75-100.