HUMAN-COMPUTER INTERACTION, 2001, Volume 16, pp. 1–38

Note-Taking for Self-Explanation and Problem Solving

J. Gregory Trafton
Naval Research Laboratory

Susan B. Trickett
George Mason University

ABSTRACT

We explore the effects of interfaces to take notes on problem solving and learning in a scientific discovery domain. In 2 experiments (1 correlational, 1 experimental), participants solved a series of 5 scientific reasoning problems in a computer environment. We provided some participants with access to an online notepad and found 3 main results: (a) Using the notepad helped participants solve the problems more accurately; (b) the benefits of using the notepad persisted after participants had stopped using it; and (c) participants who used the notepad for problem solving and self-explanation learned more, regardless of the type of notepad interface that was provided. Implications for learning systems with online notepads are discussed.

Greg Trafton is a cognitive scientist with an interest in HCI methods for improving learning environments and intelligent tutoring systems; he is a researcher at the Naval Research Laboratory. **Susan Trickett** is a doctoral student in cognitive psychology, with interests in the design of educational technology and science education; she is at George Mason University.

CONTENTS

1. INTRODUCTION

There are currently many different computer-based learning systems available, running the gamut from discovery microworlds (e.g., White, 1993) to highly structured intelligent tutoring systems (e.g., Corbett, Koedinger, & Anderson, 1997; Milson, Lewis, & Anderson, 1990). The designer of such systems has many decisions to make at many levels—from the programming language to be used to the color of the buttons the user will press.

One of the main decisions that the designer of a computer-based learning system must make concerns the system's architecture—that is, how a particular theoretical perspective is instantiated. For example, some intelligent tutoring systems use a model-tracing technique to match the learner's actions against those an expert problems solver would take (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Corbett & Anderson, 1995; Koedinger, Anderson, Hadley, & Mark, 1997; Reiser, Kimberg, Lovett, & Ranney, 1992; Trafton & Reiser, 1991; VanLehn, Niu, Siler, & Gertner, 1998). Other systems are based on the assumption that students will integrate multiple representations of the same concept, but how this integration happens differs across theories and systems (Koedinger & Anderson, 1993; Kozma, Russell, Jones, Marx, & Davis, 1996; Roschelle, 1992; Schwarz & Dreyfus, 1993; White, 1993). Decisions about the theoretical underpinnings of the system design are crucial because they shape the interactions that are possible between the user and the system. However, users themselves are often unaware of—and indifferent to—the theory or architecture that the learning system uses.

In contrast, design decisions about the interface affect the user directly because they determine how the user will interact with the system. Many of these decisions are not specific to learning environments; for example, if users try to take an illegal action, should they be alerted by sound, text, or some other means? Should they type commands, or select from a menu of options? However, a learning environment differs from other types of systems in that it should provide sufficient guidance and support for users as they learn to master the concepts under instruction. For example, there may be specific strategies that are beneficial, but not essential, for learners to use as they carry out the different learning activities. Designers can possibly improve their learning system by providing tools that support the use of these strategies.

Several effective domain-general strategies to improve learning have been identified. These strategies include self-explanation (Chi, Bassok, Lewis, Reimann, & Glaser, 1989); a cycle of prediction, testing, and reflection (Salzman, Dede, & Loftin, 1999; White, 1993); reflective inquiry (Tabak, Sandoval, Smith, Steinmuller, & Reiser, 1998); and the development of an appropriate and well-crafted “driving question” (Barron et al., 1998). Each of these strategies has been associated with successful student learning in a variety of domains and instructional environments.

In some cases, tools to support effective learning strategies have been integrated into the design of computer-based learning environments. For example, BGuILE (Biology Guided Inquiry Learning Environment) is a scientific learning environment (Loh et al., in press; Sandoval & Reiser, 1997, 1998; Tabak, Smith, Sandoval, & Reiser, 1996). Reiser and his colleagues argued that a crucial strategy for scientific inquiry in any domain is the building of explanations. They therefore included in the design of BGuILE a software tool, the ExplanationConstructor, which is designed to help students construct explanations, thereby satisfying four criteria of scientific inquiry (Sandoval & Reiser, 1997, 1998). Because the construction of explanations is a domain-general strategy, this feature of the design can be included in any scientific learning environment, regardless of the particular domain.

One strategy that may help learners as they solve problems is taking notes. Note-taking is a very general strategy that has been well studied in a number of traditional learning environments, such as classrooms and lectures. It frequently has been found to boost student performance on recall of information (e.g., Kiewra, 1985, 1989; Rinehart & Thomas, 1993; Spires, 1993; Van Meter, Yokoi, & Pressley, 1994). However, little research has been done on the effects of note-taking in problem-solving environments. Although some problem-solving environments, such as ThinkerTools (White, 1993), Smithtown (Shute & Glaser, 1991), Knowledge Integration Environment (Linn, in press; Slotta & Linn, in press), and BGuILE (Sandoval & Reiser, 1997), have included note-taking facilities, no controlled studies of their effectiveness have been conducted.

Nonetheless, there is some weak evidence from studies involving Smithtown which suggests that taking notes may be beneficial. Smithtown is an economics microworld in which students "discover" the laws of supply and demand by manipulating features of various trading markets. Shute Glaser, and Raghavan (1989) investigated the relation between specific student behaviors and learning results and found that use of Smithtown's notepad was associated with improved performance. However, Smithtown's notepad merely allowed students to copy the results of experiments they had run; it seems, therefore, to have functioned primarily as an external memory aid rather than as a means to scaffold or facilitate problem solving. Thus, whether note-taking improves performance in problem-solving environments remains an open question.

If note-taking is indeed an effective strategy for learning, one potential advantage is that it should be a straightforward matter to integrate support for this strategy into systems for learning. The question for the designer of such systems, then, becomes what kind of note-taking facility to include. The flexibility of computer-based learning environments allows for various designs of notepad, each with the potential to support a different kind of note-taking.

The notepad's design is potentially of great importance because the type of note-taking in which students engage appears to be instrumental in their performance, at least on recall tasks. Although many studies have shown that students who take notes perform better than those who do not, a number of other studies have found no advantage for students who take notes (see Kiewra, 1985, for a review). Kiewra suggested that these mixed results are due to the kind of note-taking participants practice, which may involve the student in different levels of engagement. Sometimes participants merely copy verbatim what is read or heard, involving very little engagement (e.g., Laidlaw, Skok, & McLaughlin, 1993). At other times, participants employ "conceptual note-taking" (e.g., Rickards & McCormick, 1988), summarization (e.g., King, 1992), or self-questioning (e.g., Spires, 1993). According to Kiewra, these latter types of note-taking all involve significant levels of engagement.

One interpretation of these different levels of engagement during note-taking is that some involve the students in active problem solving and self-explanation, whereas others do not. Several studies show that performance and learning improve to the extent that students engage in problem-solving activities (e.g., Anderson, Conrad, & Corbett, 1993; Trafton, 1994; Trafton & Reiser, 1993). Thus, problem solving and active engagement with a task are key components of a student's learning. According to this understanding of problem solving, copying verbatim does not involve the student in problem solving, whereas more strategic forms of note-taking, such as conceptual note-taking, summarizing, or self-questioning, require the student to perform some problem solving and self-explanation.

Viewed from a problem-solving perspective, note-taking strategies that involve some form of problem solving and self-explanation should lead to better performance than strategies that do not. The results of several note-taking studies support this interpretation. Copying, which does not engage the student in problem solving, does not result in better learning (e.g., Laidlaw et al., 1993). In fact, this kind of note-taking is no more effective than underlining (Ayer & Milson, 1993). On the other hand, "conceptual note-taking" and summarizing, which do involve problem solving and self-explanation, are more effective than merely copying material (King, 1992; Rickards & McCormick, 1988). Likewise, note-taking that involves self-questioning or reorganizing material, which also engage the student in problem solving, similarly results in better performance than copying verbatim (Shimmerlik & Nolan, 1976; Spires, 1993).

Our interpretation of the note-taking literature predicts that note-taking facilities that engage the student in self-explanation and problem solving (SEPS) will be more likely to improve performance than those that do not. Currently, however, the type of notepad that is most effective also remains an open question.

We have identified two questions for investigation. Does taking notes help performance in a problem-solving environment? If so, what kind of note-taking facility best improves problem-solving performance? Both of these questions address issues of performance within a specific environment. However, a current emphasis in education is on facilitating students' understanding at a conceptual level, not merely producing better performance (e.g., National Council of Teachers of Mathematics Commission on Standards for School Mathematics, 1989). Therefore, an important part of the success of any learning system is the extent to which learners are able to perform tasks later on, independent of the scaffolding offered by the system.

To these questions we can add a third: Is there a future benefit to the student of taking notes? That is, does taking notes during a task scaffold future efforts at performing similar tasks when the student no longer takes notes? If taking notes serves this kind of scaffolding function, students should continue to perform well on the problem-solving tasks they have been learning to solve even without the benefit of a note-taking facility because they have come to understand the task at a conceptual level.

To investigate these questions, we conducted several experiments of students performing some simple scientific reasoning tasks in an environment that provided access to a note-taking facility. We chose a scientific reasoning task because it involves a significant degree of problem solving and because mastery of the scientific method is considered crucial to the enterprise of science, insofar as it applies across scientific domains. However, many studies show that although some nonscientists do conform to a normative model of scientific reasoning, many do not (e.g., Klahr & Dunbar, 1988; Shute & Glaser, 1990; Trickett, Trafton, & Raymond, 1998; Wason & Johnson-Laird, 1972). Instead, people frequently adopt other strategies, such as conducting experiments without a hypothesis (Klahr & Dunbar, 1988) or even generating all possible experiments (Trickett et al., 1998). 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). Thus, these tasks represent the type of task for which people are likely to benefit from support in the problem-solving environment.

The first experiment presents a reanalysis of three experiments in which students had access to a freeform notepad. We analyzed the notepad entries to identify several types of note-taking strategy. We then examined the relation between these strategies and the students' performance on the scientific reasoning tasks. In particular, we wanted to discern whether note-taking was, overall, an effective strategy for problem solving and learning, and whether some note-taking strategies might be more effective than others. In the second experiment, we introduced a control group that did not have access to any type of note-taking facility, and we developed different designs of notepad to elicit different levels and types of problem solving. Our purpose was to replicate the results of the first study in an experimental fashion and to investigate the effects on problem solving and learning of the different designs of notepad.

2. EXPERIMENT 1: AN EXPLORATION OF NOTEPAD USE

Experiment 1 was an exploratory study that examined the relations among note-taking, problem solving, and learning. We conducted three separate studies 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 three studies, participants performed the same tasks and used an identical interface. In the first two studies, we had two experimental conditions that required the participant to perform either the same task repeatedly or a set of isomorphic tasks. There were only minor procedural differences among the three studies. Figure 1 summarizes relevant differences and similarities among them.

It is important to note here that none of these studies was specifically designed to investigate note-taking. The analyses and results discussed subsequently are therefore correlational and must be viewed with the usual caveats attached to this type of analysis. However, we also note that this approach addresses spontaneous rather than induced note-taking behaviors among participants and thus has the advantage that it captures their "real-world" behavior. On the other hand, a further disadvantage of this approach is that there may be few instances of overall use of the notepad within any one study. In this case, our analyses may suffer from lack of power.

To increase the power of our analyses, we therefore combined the data from all three studies under the rubric Experiment 1. Before we combined data, however, it was important to determine that there was no interacting effect of study and notepad use on the accuracy of the participants' solutions. In addition, because the three studies involved two different conditions, we also had to determine that there were no differences in use of the notepad across conditions. The following analyses address these two issues.

First, we conducted an analysis of variance (ANOVA) with overall accuracy as the dependent variable and study and use of the notepad as independent variables. The results of this analysis show that there was no main effect of study on performance accuracy, $F(2, 441) < 1$, $MSE = .2$. The interaction between study and notepad use was also nonsignificant, $F(4, 441) = 2.12$, $MSE = .45$, $p > .05$. Thus, the analysis shows that participants in one of the three studies were not more likely to do better or worse than participants in one of the other studies. Furthermore, it shows that when participants used the notepad, they were neither more nor less likely to answer the problem correctly as a function of the particular study they took part in.

Second, to ascertain that there were no quantitative differences between condition 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 nonsignificant, $F(1, 88) < 1$, $MSE = .2$. This result shows that participants in the same-task condition were neither more nor less likely to use the notepad than participants in the isomorph condition.

The results of these two analyses indicate that there were no significant performance differences among our participants based on the study they participated in and that there were no significant differences among note-taking behaviors based on the experimental condition to which they were assigned. Consequently, we combined the data from these three studies to increase our analytical power.

Figure 1. **Similarities and differences between studies in Experiment 1.**

	Study 1	**Study 2**	**Study 3**
Semester	Fall 1997	Fall 1997	Spring 1998
Conditions	Same-task/isomorph	Same-task/isomorph	Isomorph only
Verbal protocols	Yes	No	No
Number of tasks	5	5	5
Access to notepad	Yes	Yes	Yes

2.1. Method

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 men and 48 women).[1]

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 three 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 result was displayed in a text box that remained visible throughout the task. If the text box became full (after 10 experiments were run), participants could use a scroll bar to view the results of their earlier experiments.

---

[1] Although we did not collect demographic data regarding our participants' age, we should note that George Mason University undergraduates are more diverse than many undergraduate sample populations in terms of age, Scholastic Assessment Test scores, and other factors. Thus, our sample included several older participants.

The interface also included a notepad, consisting of a blank text box, on which participants could enter information or comments if they chose. Figure 2 shows a screen snapshot of the interface for the roller coaster task (note, however, that it illustrates a four-variable version of the task, which was used in Experiment 2). 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 to identify that variable.

A task analysis 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). Finally, they could find two different experiments that yielded the same result and deduce that the variable that had a different setting in this pair of experiments must not affect the ride (e.g., LEFT, LEFT, LEFT: looping, upside-down ride; LEFT, LEFT, RIGHT: looping, upside-down ride).

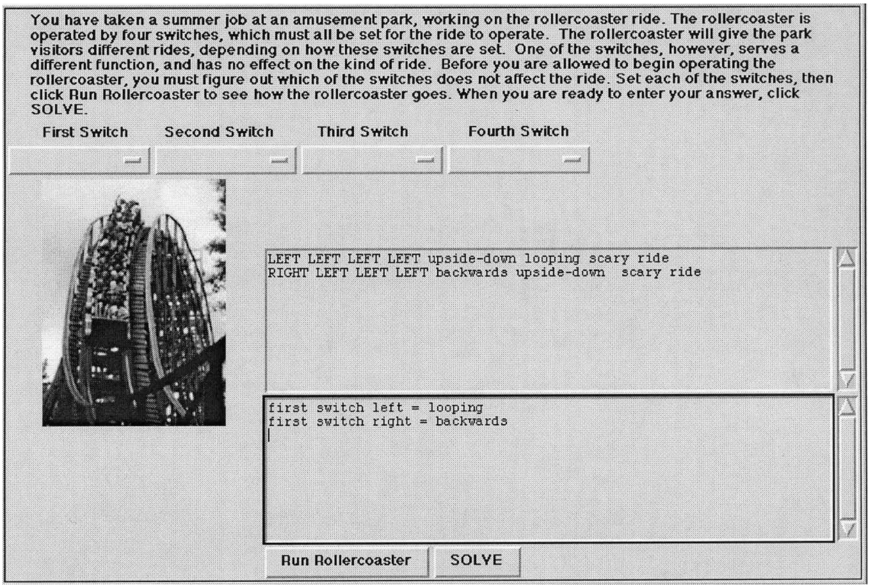
Design

There were five different tasks, as described previously. In Studies 1 and 2, there were two conditions: a same-task condition and an isomorph 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; that is, all were in the isomorph condition. In both conditions, the correct solution for a task was generated randomly 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 proto-

Figure 2. **Screen snapshot of roller coaster task. The upper text box contains the results of the experiments after the Run Roller Coaster button was pushed; the lower text box was a freeform notepad.**



You have taken a summer job at an amusement park, working on the rollercoaster ride. The rollercoaster is operated by four switches, which must all be set for the ride to operate. The rollercoaster will give the park visitors different rides, depending on how these switches are set. One of the switches, however, serves a different function, and has no effect on the kind of ride. Before you are allowed to begin operating the rollercoaster, you must figure out which of the switches does not affect the ride. Set each of the switches, then click Run Rollercoaster to see how the rollercoaster goes. When you are ready to enter your answer, click SOLVE.

```
LEFT LEFT LEFT LEFT upside-down looping scary ride
RIGHT LEFT LEFT LEFT backwards upside-down  scary ride
```

```
first switch left = looping
first switch right = backwards
```

cols were collected. We used keystroke data to determine the accuracy of each participant's solution for each task. We counted as correct only those tasks for which the data necessary to determining the correct solution were available (i.e., the participant had run a sufficient number of experiments). This procedure reduced, but did not entirely eliminate, the possibility that an answer would be correct as a result of chance. This measure showed that participants guessed the answer correctly on fewer than 5% of the tasks.

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 participants made an entry on the notepad was coded as a use of the notepad. We were also interested in any possible carryover effect of using the notepad. Consequently, after participants 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 post-use of the notepad. All other tasks on which participants 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). Figure 3 illustrates this coding scheme.

Figure 3. **Example of how notepad use was coded.**

Participant ID	Task	Did Participant Use Notepad?	Notepad Code
37	1	No	Non-use
37	2	No	Non-use
37	3	Yes	Use
37	4	Yes	Use
37	5	No	Post-use

In addition to coding these stages of notepad use, we coded how participants used the notepad. Several types of notepad entry were identified, as follows. Some participants identified what the variables did (Identify Variable). This category includes identifying the effect of each level of a variable or 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 related to the strategies previously outlined by which participants could successfully solve the tasks. Some participants used the notepad to rerepresent the problem space or to rerepresent the experiments they had run (Rerepresentation). Some participants used the notepad to state a hypothesis (State Hypothesis). A few participants used the notepad to identify whether a variable had an effect, without specifying the effect (Effect); to keep track of the experiments they had run (Keep Track); or to copy text that was already visible on screen (Copy).

In some cases, participants wrote notes that appeared to engage the task but indicated that these participants had the wrong representation of the task. For example, although the task required that participants identify the variable that had no effect, some participants took an engineering approach (Schauble, Klopfer, & Raghavan, 1991; Trickett et al., 1998). They attempted to find the combination of variables that brought about what they considered a desirable state (e.g., make the roller coaster go very fast). Entries in the notepad that reflected a wrong representation were coded separately (Wrong Representation). Notepad entries that demonstrated more than one category of use were coded as “Mixed” entries.

Six entries on the notepad did not fit any of the categories just described. 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. Figure 4 summarizes the coding scheme for types of entry and gives examples of each category.

Figure 4. **Codes and examples (from the data) for type of notepad use.**

Type of Use	Example
Identify Variable	3rd switch L = very fast; 3rd R = upside down; 2nd switch affects looping or backwards
2-Same	RRR and RRL same ride; LLL and LLR same ride
Rerepresent	1 + 1 + 1 = very fast backwards; 1 + 2 + 1 = very fast looping
State Hypothesis	It may be the second switch
Effect	1st switch L and 1st switch R don't affect; 2nd switch does affect
Keep Track	LLL; LRL; RRR; RLR; LLR; RLL
Copy	LRL very fast backwards ride; LRL very fast looping ride
Mixed	First left = looping or backwards; third switch is a possibility
Wrong Representation	LLR makes it very fast; LRR goes very fast
Discard (recoded as non-use)	If the answer is not chromosome 5 then these surveys should be deemed ineffective due to the fact that the way a possible answer is worded should not be meant to trick the test taker.

Note. L = left; R = right.

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 could do so if they wished; they were not told what to write on the notepad.

2.2. Results and Discussion

We first analyzed the extent to which participants used the notepad in each of the three studies. There were 90 participants in the three studies; 39 (43%) of these participants used the notepad at least once. In each study, there were 30 participants, who each performed 5 tasks (i.e., there were 150 tasks or opportunities for notepad use in each study). 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 44 uses (29% of tasks). In all three studies combined, there were 86 uses of the notepad over 450 tasks (19% of tasks).

Relation Between Notepad Use and Performance

Our first question was whether using the notepad helped participants solve the problems correctly. We expected that participants who used the notepad would perform better on these problems than those who did not. However, in the three studies that compose Experiment 1, there was no control group of participants who did not have the opportunity to use the notepad. Thus, it was not appropriate to make general comparisons between the performance of those participants who used the notepad and those who did not. We could not, for example, compare the overall performance on the series of five tasks of those who used the notepad with the overall performance of those who did not because we did not control the task on which the notepad users first started to use the notepad. Participants who used the notepad did not always do so on the first task. Of the 39 participants who used the notepad, 16 (41%) used it for the first time on the second, third, fourth, or even fifth task. Even a task-by-task comparison was not appropriate because in general performance improved as participants progressed through the series of tasks (Trickett et al., 1998). Consequently, any effect of using the notepad on the later tasks would be confounded by this overall learning.

To investigate whether participants who used the notepad performed better than those who did not, however, we could compare performance of the two groups on the very first task. Such a comparison may be particularly informative because, in general, participants found the task very difficult and did rather poorly on the first task, with only about half the participants (57%) solving the first problem correctly. Of the 39 participants who used the notepad, 23 used it on the first task. Of those 23 participants, 18 (78%) solved the problem correctly. Of the remaining 67 participants who did not use the notepad on the first task, 33 (49%) solved the problem correctly. Figure 5 summarizes these results. A chi-square test on these data was significant, $\chi^2(1, N = \text{X}) = 5.87, p < .05$.

This result shows that on the first task, where a direct comparison can be made, participants who used the notepad were correct significantly more frequently than those who did not use it. Furthermore, because on the first task all participants were otherwise "on an equal footing," the analysis strongly suggests that using the notepad was helpful to participants and increased the likelihood that they would solve the problem correctly. Of course, it is possible that participants who used the notepad differed in some way that caused them to use the notepad, but this argument is addressed in the summary of Experiment 1, as well as in Experiment 2.

Another approach to determining the relation between the use of the notepad and participants' performance on the tasks is to examine this relation among the "notepad users" only—that is, among that subset of 39 participants

***Figure 5.* Performance of notepad users and non-users on Task 1 (Experiment 1).**

	Correct	Incorrect
Used notepad	18 (78%)	5 (22%)
Did not use notepad	33 (49%)	34 (51%)

who at any point chose to use the notepad. As we discussed previously, not all participants who used the notepad did so for the first time on the first task; consequently, we can examine the performance of the notepad users on tasks for which they did not use the notepad (non-use) and on tasks for which they did use it (use).

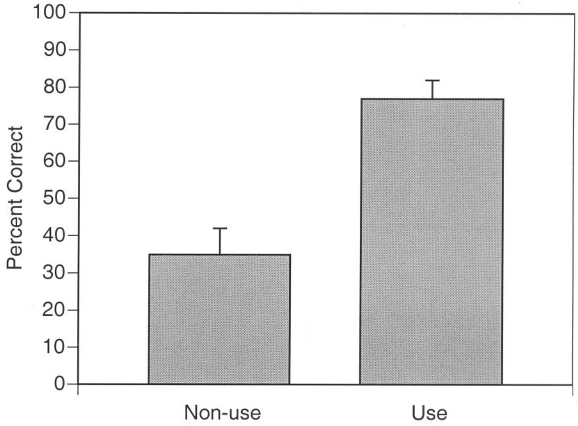
We investigated the accuracy of the notepad users' solutions at both these stages. Figure 6 summarizes these results. An ANOVA comparing percentage correct at both stages of notepad use (non-use and use) was significant, $F(1, 127) = 25.4$, $MSE = .2$, $p < .01$, with means of .35 and .77, respectively. This analysis shows that among those participants who used the notepad on at least one task, performance was significantly better when they used it than when they did not use it. Because in general participants' performance improved across the sequence of five tasks, we also performed an analysis of covariance (ANCOVA) with task as the covariate, to control for the effect of learning over time. The results of the ANCOVA were significant, $F(1, 126) = 25.8$, $p < .01$. Thus, the results show that participants who used the notepad performed better, even accounting for the fact that, in general, performance improved later in the sequence of tasks. Together, these analyses suggest, although they do not conclusively show, that using the notepad did indeed help participants' performance on these problem-solving tasks.

How Participants Used the Notepad

The analyses described previously show both that participants who used the notepad on the first task performed better than those who did not and that among those participants who used the notepad, performance was better when they used it than when they did not. However, the analyses do not differentiate among different types of notepad use. Our second question was, What kind of use did participants make of the notepad? We now address this question and further investigate whether there is a relation between the type of use to which the notepad is put and participants' performance on the task.

To investigate these issues, we first categorized the type of use that the participants made of the notepad, according to the coding scheme discussed previously, and then examined the type of notepad use in relation to the accuracy

Figure 6. **Percentage correct for use and non-use among notepad users (Experiment 1). Error bars are standard error of the mean.**

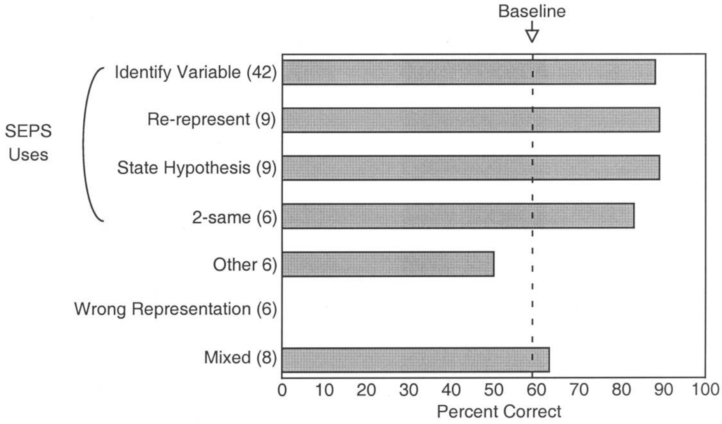


of the solution. Some types of notepad entry were very sparsely represented. Those categories for which there were three or fewer instances were combined under the notation Other. These categories included identifying whether a variable had an effect, keeping track of experiments, and copying. Because the number of notepad entries for several categories was rather 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. Figure 7 shows these percentages, together with the actual number of instances of each type of entry.

The accuracy rate among all participants on tasks for which they did not use the notepad was 59%. We refer to this as the baseline. Four types of entry on the notepad (Identify Variable, Rerepresent, State Hypothesis, and 2-Same) were at least 24% above baseline. Entries on the notepad that made mixed use of the notepad were approximately the same as the baseline. Types of entry coded as Other were below baseline.

Thus, it appears 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, rerepresent the problem or the experiments run, or state a hypothesis seems related to successful performance. Why might these uses of the noted be helpful as participants tried to solve the tasks? There are several possible explanations. 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. Identifying what a variable does and identi-

***Figure 7.* Percentage correct for different types of notepad use (Experiment 1). Dashed line represents baseline.**



fying two identical results correspond to strategies by which the tasks could be solved. However, a more general explanation is that both rerepresentation and stating a hypothesis, as well as the strategic uses, engage the participant in problem solving and self-explanation. Recall that our interpretation of the mixed results of the note-taking literature indicated that note-taking would facilitate performance insofar as it engaged the participant in problem solving. These results further support the suggestion that taking notes improved performance because it facilitated problem solving and self-explanation.

Several observations can be made concerning the uses of the notepad that were not associated with more correct responses than the baseline. The uses coded as Other included simply copying data that were already visible on the screen. This use of the notepad did not engage the participants in any problem solving and therefore would not be likely to be associated with good performance. In six other cases, the entries on the notepad indicated that the participant had a wrong representation of the task—for example, understood that the objective was to achieve a certain goal state rather than determine the respective effects of the different variables. Although the entries on the notepad appeared to engage the task, at least as the participant interpreted it, in every single case the participant failed to solve the problem correctly. Thus, it appears that if the participant set off on the wrong track, using the notepad was not helpful. This suggests that if the use of the notepad was directed toward exploring that wrong track, rather than building alternative representations of the task, performance would not improve.

As for the mixed use of the notepad, participants were correct in five of the eight cases. Mixed use of the notepad could involve two (or more) of the uses identified as helpful, two (or more) of the uses identified as not helpful, or some combination of helpful and unhelpful strategies. To the extent that the final use of the notepad was helpful, we might expect the mixed use of the notepad to be helpful overall, insofar as the final use might indicate a shift in the participants' problem-solving strategy. Indeed, in six of the eight cases, the final use was Identify Variable, 2-Same, or State Hypothesis. Participants were correct on four of these six tasks. For the two tasks on which they were incorrect, the earlier use was Wrong Representation; again, although the participants attempted to engage the task, they were unable to use the notepad fruitfully because their representation of the task was incorrect.

Scaffolding Effect of the Notepad

It appears, then, that using the notepad could be helpful in these problem-solving tasks. Furthermore, uses of the notepad that guided participants toward a good solution strategy or that otherwise engaged them in problem solving were more likely associated with successful performance than those uses that did not engage the participants in problem solving. We believe that the notepad itself was a scaffold for the students. Having the online notepad available allowed the students to perform more effective problem solving by using the notepad. After the students stopped using the notepad, however, the scaffold was not being used, and perhaps the advantages of that scaffold disappeared. It could be, however, that the advantages of the notepad persisted even after the students stopped using the notepad. If the advantages did persist, there would be a scaffolding effect of the notepad, which would be a kind of "cognitive residue" left over from using the scaffold. Our third question was thus whether there was a scaffolding, or carryover, effect of using the notepad.

To investigate this question, we compared performance among the notepad users on those tasks coded as post-use with their performance on tasks coded as non-use and use. Recall that post-use refers to those tasks on which participants did not use the notepad but had used it previously. Thus, if there was a benefit to using the notepad that persisted beyond the immediate task on which it was used, we would expect to find superior performance on these post-use tasks.

An ANOVA comparing performance at these three stages of notepad use (non-use, use, and post-use) was significant, $F(2, 192) = 18.15$, $MSE = .18$, $p < .01$, with means of .35, .77, and .82, respectively.[2] A Scheffé test indicated that

---

[2] Again, because in general, performance improved later in the sequence of tasks, and because post-use tasks occurred later in the sequence than either non-use or use, we also performed an ANCOVA, with task as covariate. This ANCOVA was significant, $F(2, 192) = 17.19$, $p < .01$.

there was a significant difference between non-use on one hand and use and post-use on the other hand ($p < .05$). Figure 8 summarizes these results.

This analysis points to a carryover effect of using the notepad. Most participants (69%) who used the notepad did not continue to use it on all the remaining tasks but rather stopped using it after one or more uses. It appears that, in these cases, participants did as well as, or even slightly better than, they did on tasks for which they were actually using the notepad. Thus, the notepad appears to have served as a scaffold for participants' problem solving. Once the scaffold was removed (by the participants' own choice), they were able to maintain their performance at the same level without the benefit of the scaffolding. This result thus shows an association between using the notepad and improved learning on these tasks.

2.3. Summary of Experiment 1

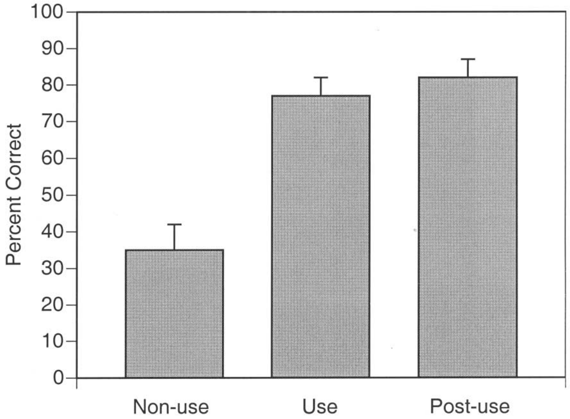
The results described here show that participants who used the notepad were more likely to solve the problem correctly than those who did not use it, at least on the first task where comparison is possible. The results show that among those participants who used the notepad on at least one task, performance was significantly better when they used it than when they did not use it. However, only some uses of the notepad were associated with achieving the correct solution; all these uses engaged the participant in some level of SEPS. Finally, the benefits of using the notepad were maintained even after participants stopped using the notepad on subsequent tasks; thus, the notepad seems to have had facilitated participants' problem solving.

However, these conclusions must be considered tentative because of the correlational nature of the data. We cannot decisively reject other possible explanation, such as that the notepad users were "better" problem solvers or were for some reason intrinsically inclined to use the notepad. Furthermore, the results do not inform us about the most effective design for an online notepad. Consequently, our next step was to test these results experimentally by manipulating use of the notepad and providing the appropriate controls and to investigate the effect of notepad design on performance.

3. EXPERIMENT 2: A STUDY OF DIFFERENT NOTEPADS

As discussed previously, Experiment 2 was designed to formally test several hypotheses that arose from the exploratory investigations described in Experiment 1. First, we predicted that participants who used a notepad to take notes during problem solving would perform better than participants who did not take notes. Second, we predicted that participants who used a notepad

Figure 8. **Percentage correct for non-use, use, and post-use among notepad users (Experiment 1). Error bars are standard error of the mean.**



would continue to outperform those who did not, even after they no longer had access to the notepad.

The results of Experiment 1 strongly suggest that a notepad that engages participants in SEPS will lead to better performance than a notepad that does not require problem solving. The interface provided participants with a notepad that allowed a variety of uses but did not specifically guide them toward any particular kind of use. However, as Figure 7 shows, one of the most frequent and most effective uses of the notepad was to figure out the effect of each variable. Recall that this was one of the strategies by which the task could be solved. It would seem, then, that providing participants with a notepad that guides them to use this strategy would be a highly effective means of improving performance on the tasks. Insofar as participants use such a notepad to support this strategy, we would expect them to perform better than participants who are given a notepad that does not constrain their use. Experiment 2 thus also investigated the effect of the notepad design on problem-solving performance and learning.

3.1. Method

Participants

Participants were George Mason University undergraduates who received course credit for their participation. There were 58 participants in this experiment (42 women and 16 men). (An additional participant was excluded from the analysis because technical difficulties resulted in only partial data collection.) Participants were assigned randomly to one of four conditions. In three of these conditions (freeform notepad, semistructured notepad, and structured notepad), participants had access to an online notepad. (The three different designs of notepad are described subsequently.) The fourth condition was a control group that did not have access to a notepad as they performed the tasks. There were 14 participants in the freeform notepad group, 14 in the semistructured notepad group, 15 in the structured notepad group, and 15 in the control group.

Materials

Participants performed the same five tasks as participants in Experiment 1. However, to reduce the possibility that participants would choose the correct answer by chance, we created another variable so that each task involved four variables rather than the original three. In addition, we changed the interface to investigate the effect of notepads involving different levels of problem solving.

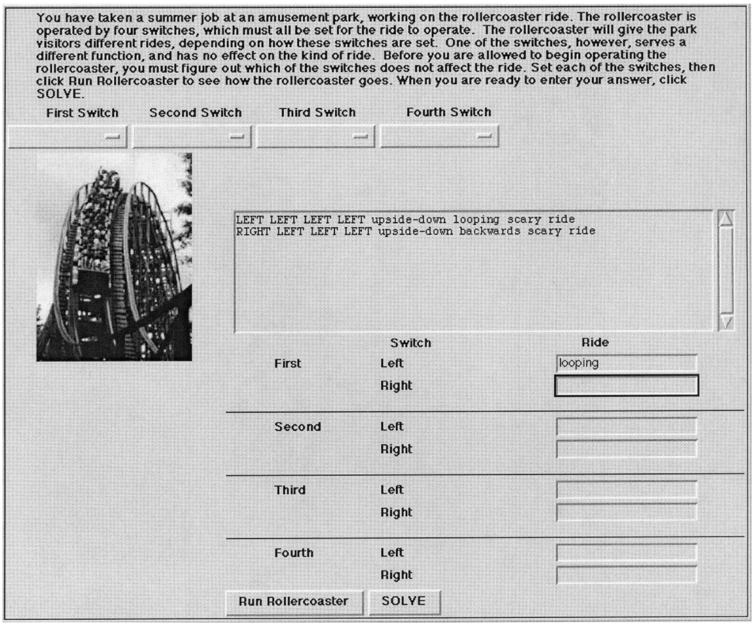
In one condition, the notepad was identical to the one described in Experiment 1; that is, it was a blank text box in which participants could type anything they wanted (freeform notepad). Figure 2 shows a screen snapshot of this version of the interface. This notepad did not guide the participants in specific ways toward the problem solution.

In the second condition, the notepad consisted of a listing of each level of all the variables that participants could manipulate. Adjacent to each variable was a blank text box into which participants could enter their findings about that variable's effect (semistructured notepad). Figure 9 shows a screen snapshot of this version of the interface. This notepad guided the participants to think about the problem in terms of each level of each variable; however, participants still had to determine what to type into the blank text box.

In the third condition, the notepad also consisted of a listing of each level of all the variables, as in the semistructured notepad condition. However, adjacent to each variable was a button that opened a pop-up menu listing all the possible effects (including no effect) that the variable could have (structured notepad). Figure 10 shows a screen snapshot of this version of the interface. Thus, participants in this condition merely had to select from a predetermined list which effect they thought the variable had. Consequently, the structure of the notepad guided them in a specific way to the problem solution.

The fourth condition was a control group in which participants had no notepad. In all other respects, the interface was identical to the interfaces for the other three conditions. In all three notepad conditions, the notepad was available for only the first two tasks in the series of five tasks. On the third, fourth, and fifth tasks, there was no notepad, and the interface in each became identical to that used by the no-notepad group. Making the notepad unavail-

Figure 9. **Screen snapshot of interface for semistructured notepad.**

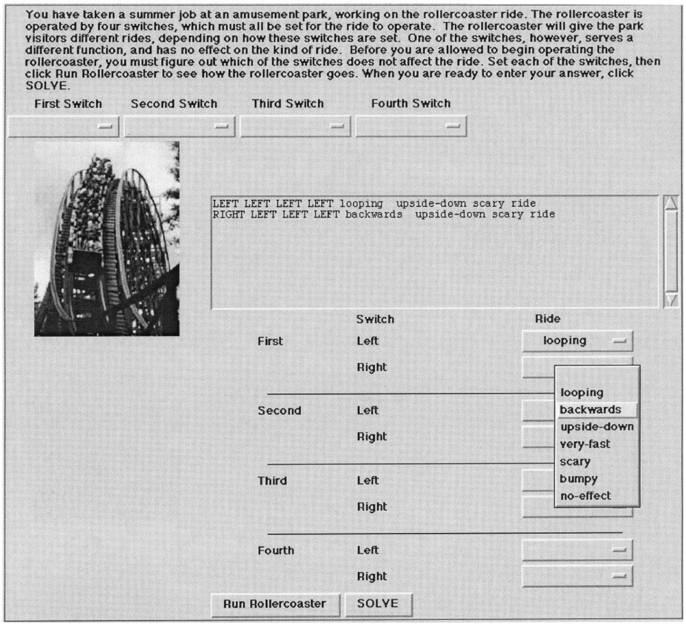


You have taken a summer job at an amusement park, working on the rollercoaster ride. The rollercoaster is operated by four switches, which must all be set for the ride to operate. The rollercoaster will give the park visitors different rides, depending on how these switches are set. One of the switches, however, serves a different function, and has no effect on the kind of ride. Before you are allowed to begin operating the rollercoaster, you must figure out which of the switches does not affect the ride. Set each of the switches, then click Run Rollercoaster to see how the rollercoaster goes. When you are ready to enter your answer, click SOLVE.

```
LEFT LEFT LEFT LEFT upside-down looping scary ride
RIGHT LEFT LEFT LEFT upside-down backwards scary ride
```

	Switch	Ride
First	Left	looping
	Right	
Second	Left	
	Right	
Third	Left	
	Right	
Fourth	Left	
	Right	

Figure 10. **Screen snapshot of interface for structured notepad.**



You have taken a summer job at an amusement park, working on the rollercoaster ride. The rollercoaster is operated by four switches, which must all be set for the ride to operate. The rollercoaster will give the park visitors different rides, depending on how these switches are set. One of the switches, however, serves a different function, and has no effect on the kind of ride. Before you are allowed to begin operating the rollercoaster, you must figure out which of the switches does not affect the ride. Set each of the switches, then click Run Rollercoaster to see how the rollercoaster goes. When you are ready to enter your answer, click SOLVE.

```
LEFT LEFT LEFT LEFT looping  upside-down scary ride
RIGHT LEFT LEFT LEFT backwards  upside-down scary ride
```

	Switch	Ride
First	Left	looping
	Right	
Second	Left	
	Right	
Third	Left	
	Right	
Fourth	Left	
	Right	

able in this way allowed us to control when participants stopped using the notepad. Thus, we were able to test the prediction that participants who used the notepad would continue to outperform those who did not, even after they no longer had access to the notepad.

How might the different designs of notepad affect problem-solving strategies? On one hand, one might predict that the freeform notepad would engage participants in the greatest level of SEPS, whereas the structured notepad would guide participants to the correct solution and thus engage them in less SEPS. According to this interpretation, the semistructured notepad might be seen as engaging participants in some intermediate level of problem solving. If this were the case, we would expect that participants using the freeform notepad would demonstrate superior performance, with participants in the structured notepad condition showing worse performance and participants in the semistructured notepad condition falling somewhere in between.

Alternatively, though, one might predict that the scaffolding provided by the structured notepad would guide the students initially so that they would develop a good strategy to solve the problems. Use of such a strategy then might continue after the notepad was no longer available, leading to overall better performance and learning. According to this view, we would expect that the participants in the freeform notepad condition, lacking such scaffolding, would demonstrate less effective performance and learning. In this case, we also would expect the participants in the semistructured notepad condition to fall somewhere in between, because the notepad available to them provided an intermediate level of scaffolding.

We planned to engage issues surrounding the effect of the notepad design on performance by analyzing the way participants used the notepad, the relation between type of notepad use and problem solving, and participants' performance on the task.

Design

This experiment was a single-factor (notepad type), between-subject design. There were four levels of notepad type, resulting in four conditions, as described previously: freeform notepad, semistructured notepad, structured notepad, and no-notepad (the control group). Participants performed five different but isomorphic tasks, as described in Experiment 1.

Measures

Keystroke data, including entries participants made on the notepad, were collected as participants solved the tasks. We used keystroke data to determine the accuracy of each participant's solution for each task. As in Experiment 1, we coded how each participant in the notepad conditions used the notepad. These codes are described subsequently in the Results and Discussion section.

Procedure

Participants were trained on the interface appropriate to their experimental condition. They practiced designing and running experiments, viewing the results, and using the notepad (notepad conditions only). Participants in the notepad conditions were told that the notepad would be available for some but not all tasks. They were instructed to use the notepad when it was available, although they were not told specifically how they should use it. To make sure participants did use the notepad, a reminder appeared on the computer screen during the first task.

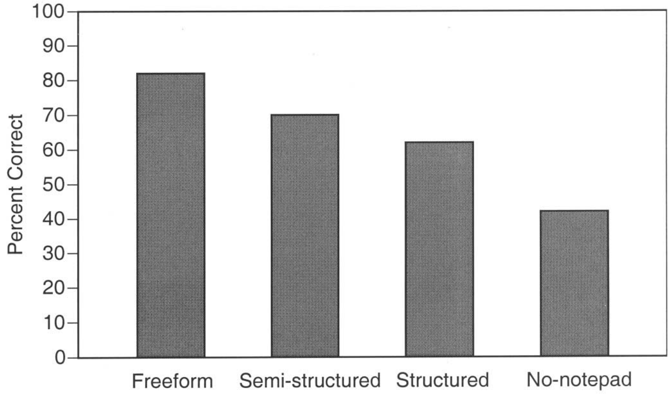
3.2. Results and Discussion

First, we determined whether participants in the three notepad conditions did use the notepad. All 43 participants in these conditions used the notepad at least once as they performed the first two tasks. Forty-one participants (95%) used the notepad on the first task, and 35 (81%) used it on the second task. The 2 participants who did not use the notepad on the first task did use it on the second task. Of the 8 participants who stopped using the notepad on the second task, 2 were in the freeform notepad group, 3 in the semistructured notepad group, and 3 in the structured notepad group. Thus, by far the majority of participants used the notepad on both tasks, as they were instructed to do, and the few who stopped using it after only one task were spread equally across conditions.

Relation Between Notepad Use and Overall Performance

Our first prediction was that participants in the notepad conditions would solve more tasks correctly than those in the control group. To test this prediction, we first performed an ANOVA on the percentage of correct answers for each group to establish that our manipulation had an overall effect. The results of this ANOVA were significant, $F(3, 54) = 3.48$, $MSE = 2.78$, $p < .05$. Mean number of correct responses for each condition were 4.1 (freeform notepad), 3.5 (semistructured notepad), 3.1 (structured notepad), and 2.1 (no-notepad). Figure 11 summarizes these results. We then performed a planned comparison to determine whether the three notepad conditions differed significantly from the no-notepad condition. The results of the comparison were significant, $T(54) = 2.87$, $p < .01$. These results support our first prediction, that participants who take notes while doing these problem-solving tasks perform better than those who do not.

Figure 11. **Mean overall performance (Experiment 2).**



Scaffolding Effect of Using the Notepad

Thus, taking notes appears to be helpful as a strategy in these tasks, but is there a continued benefit to note-taking that persists after participants cease to use the notepad? Recall that in Experiment 1 we considered participants' use of the notepad in two stages—that is, in terms of use and post-use. However, in Experiment 1 we were not able to control the task on which notepad users stopped using the notepad. In Experiment 2, we manipulated the post-use stage by eliminating the notepad from the interface on the third task for the participants in all the notepad conditions. Thus, on the first two tasks, all these participants had used the notepad, whereas on the third, fourth, and fifth tasks, none of them used the notepad. To investigate the scaffolding effect of the notepad, we conducted two separate ANOVAs on the participants' accuracy on Tasks 1 and 2 (notepad use tasks) and on Tasks 3, 4, and 5 (the post-use tasks).

The mean number of correct solutions on Tasks 1 and 2 (the notepad use tasks) was 1.4 (freeform), 1.2 (semistructured), 1.1 (structured), and 0.7 (no-notepad). Although the overall ANOVA was not significant, a planned comparison between the three notepad conditions and the control group was significant, $T(54) = 1.99$, $p = .05$. Thus, the participants in the three notepad conditions did perform better than those in the control group on those tasks for which they used the notepad. The second ANOVA, on Tasks 3, 4, and 5 (post-use), addressed the scaffolding effect—if the benefit of using the notepad carried over after participants no longer used it, they would continue to maintain this superior performance on these later tasks. The result of this ANOVA was significant, $F(\text{x, xx}) = 4.34$, $MSE = 1.16$, $p < .01$. The mean number of correct solutions on these three tasks was 2.7 (freeform), 2.3 (semistructured), 1.9

(structured), and 1.3 (no-notepad). We also performed a planned comparison to determine whether the three notepad conditions differed from the control group on these post-use tasks. The comparison was significant, $T(54) = 2.95$, $p < .01$. Figure 12 summarizes the results of these analyses.

Benefits of Using the Notepad

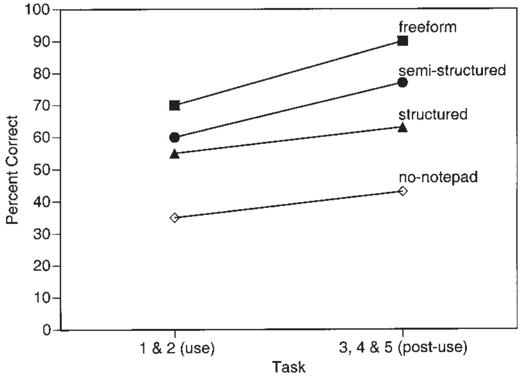
Together, these analyses show a consistent benefit to participants of taking notes during these problem-solving tasks. First, there was an overall benefit of taking notes, in terms of overall performance over the series of five tasks. Second, note-taking was beneficial in the early stages of learning to solve these problems, when the notepad was available. Finally, the benefit of taking notes persisted, and became even more pronounced, after the notepad was no longer available, so that on the last three tasks the participants in the notepad conditions continued to outperform those in the control group. Figure 12 shows this learning trend quite clearly. Figure 12 shows that participants in the notepad conditions had an immediate advantage on the first two tasks over those in the control group. On the last three tasks, when the notepad was no longer even available to them, the participants in the notepad conditions performed significantly better than those in the control group. The participants in the no-notepad control group, however, continued to struggle with the task even in the later stages, achieving an accuracy rate of only 44%. On these tasks, participants who engaged in note-taking performed better and learned more than those who did not have the opportunity to take notes.

Relation Between Notepad Design and Performance

The third issue we investigated in this study was the relation between task performance and the design of the notepad. Was one kind of notepad more likely to lead to superior problem-solving performance than another? To investigate this question, we compared performance among participants in the three notepad conditions at the two stages of notepad use previously identified.

We conducted two separate linear trend analyses, the first on the tasks for which the notepad participants actually used the notepad and the second on those tasks for which the notepad was no longer available. The results of the first analysis were nonsignificant; however, as Figure 12 shows, there was a trend in the performance of the different groups: The participants in the free-form notepad condition performed better than those in either the semistructured or the structured notepad conditions. The results of the second analysis, on the post-use stage of notepad use, were significant, $F(1, 40) = 5.1$, $MSE = .99$, $p < .05$. This result confirms the trend indicated by Figure 12: On those tasks for which the notepad was no longer available, participants in the

Figure 12. **Percentage correct for notepad use and post-use tasks (Experiment 2). Note that the mean number of correct solutions has been converted to percentage correct.**



freeform notepad condition performed best, followed by those in the semistructured notepad condition, followed by those in the structured notepad condition. Thus, the effect of the particular design of notepad appears to have become more pronounced after the notepad was no longer available to the participants.

Relation Between SEPS and Performance

We predicted that taking notes would be more likely to lead to better performance insofar as it engaged the participants in SEPS. Therefore, it is possible that the freeform notepad elicited superior performance because it engaged the participants in the highest level of SEPS. To investigate the relation between SEPS and the design of the notepad, we coded the note-taking strategy of each participant on each task for which the notepad was available. In the freeform notepad condition, we used the same codes as described in Study 1. As in Experiment 1, some Experiment 2 participants used more than one strategy. Individual strategies were identified, and the overall strategy use was coded as Mixed. Refer to Figure 4 for the codes used for the freeform notepad.

Recall that the structured and semistructured notepads were designed to guide the participants toward the Identify Variable strategy. However, not all participants in these conditions used this strategy. Some attempted to do so, with varying degrees of success. Some participants used these notepads to identify correctly the effect of each level of every variable (All-Correct Match). Some participants used this strategy but did not complete the notepad for every variable, leaving some entries blank. Those entries that they did complete, however, were correct (Incomplete–Correct Match). Some participants apparently attempted to use this strategy but made some errors in assigning effects to variables, although other variables were correctly associated with their effect (Correct–Incorrect Match).

Other participants apparently were confused about how use the notepad, and their confusion showed in two ways. First, some participants made a linear match between the result of the experiment they had run and the entries in the notepad. For example, a participant following this strategy might run an experiment in the roller coaster task in which all the switches were set to the left. The result of this experiment might be LEFT, LEFT, LEFT, LEFT, very-fast, upside-down, backward ride. In the Linear Match strategy, the participant matched very-fast with the first switch, upside-down with the second switch, and backward with the third switch. There was no obvious effect to link to the fourth variable, and so the participant simply assigned the value "no-effect." This strategy is illustrated in Figure 13. Yet other participants filled in the notepad but without attempting to isolate the effect of individual variables. For example, a participant might match the same effect to several, or even all four, variables. This strategy was coded as a Random Match.

Finally, in the semistructured notepad condition, participants could type whatever they chose in the space provided next to each variable. A few participants copied the entire result into this space (Copy), and one participant used the space to note whether the variable had an effect (Effect). As in the freeform condition, some participants used more than one strategy over the course of a task; these individual strategies were identified, and the overall strategy was coded as Mixed use of the notepad.

First, we investigated how participants in the different conditions used the notepad. Interestingly, in contrast to Experiment 1, in the freeform condition in Experiment 2 there were no uses of the notepad that reflected a wrong representation of the task or that were irrelevant to the task. Figure 14 shows the number of participants who used each strategy. Note that Figure 14 summarizes the overall use; several participants used the notepad for more than one strategy on a task, as reflected in the Mixed category. Similarly, Figure 15 shows the number of participants using each strategy in the semistructured and structured notepad conditions.

Next, we determined the extent to which each individual note-taking strategy engaged the participants in problem solving, and we assigned each overall strategy a value from 0 to 1, accordingly. In the freeform notepad condition, simply copying text that was visible on the screen was considered a non-problem-solving use of the notepad and was assigned a value of 0. All other single uses (i.e., not Mixed uses) engaged the participants in some form of problem solving and were thus considered problem-solving uses of the notepad and assigned a value of 1. To assign a value for the Mixed uses, we identified the individual strategies that made up each Mixed use. If both uses were problem-solving uses (e.g., Effect followed by Identify Variable), we consid-

***Figure 13.* Linear match strategy used by participants on the semistructured and structured notepads.**

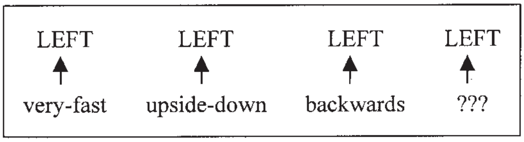


***Figure 14.* Strategy use in Experiment 2—freeform notepad.**

Strategy Use	No. of Participants Using Strategy Overall
Identify Variable	4
Rerepresent	2
State Hypothesis	1
2-Same	2
Keep Track	1
Effect	0
Copy	6
Mixed	10

***Figure 15.* Strategy use in Experiment 2—semistructured and structured notepads.**

Strategy Use	Semistructured: Participants Using Strategy Overall	Structured: Participants Using Strategy Overall
All Correct	6	7
Correct–Incorrect	1	4
Incomplete Correct	4	1
Linear	1	4
Random	4	5
Effect	1	NA
Copy	1	NA
Mixed	5	6

Note. NA = not applicable.

ered it an overall problem-solving use of the notepad and assigned it a value of 1. If a non-problem-solving use was followed by a problem-solving use (e.g., Copy followed by Identify Variable), we considered that the problem-solving use was the last use and thus assigned a value of 1 to the overall notepad use. If a problem-solving use came first and was followed by a non-problem-solving use (e.g., Identify Variable followed by Copy), we considered that there was some value to the problem-solving use but that the problem-solving use was the last use. We assigned such overall uses of the notepad a value of 0.5. However, there were only two instances of this type of SEPS followed by non-SEPS Mixed use.

In the semistructured and structured notepad conditions, the Linear Match strategy was considered a non-problem-solving use of the notepad because participants relied exclusively on a perceptual strategy. The Random strategy was also considered a non-problem-solving use of the notepad. In the semistructured notepad condition, the Copy strategy was considered a non-problem-solving use of the notepad, as it was in the freeform notepad condition. These three strategies thus were assigned a value of 0. The other individual strategies were considered problem-solving uses because the participants appeared to be attempting to match the variables to their effects (Identify Variable) along something other than perceptual or random criteria, regardless of whether they did so correctly. Thus the All-Correct, Incomplete–Correct, and Correct–Incorrect strategies were assigned a value of 1. Mixed uses of the notepad that involved two non-problem-solving uses (i.e., Random and Linear matches) were assigned a value of 0; mixed uses for which the final use was a problem-solving use were assigned a value of 1. There were no instances of SEPS followed by non-SEPS Mixed use in these conditions.

We predicted that better performance on the tasks would be associated with the SEPS uses of the notepad rather than non-SEPS uses. Using the values assigned to each overall use of the notepad, we correlated notepad use with task accuracy on Tasks 1 and 2 (the tasks for which participants had access to the notepad). The results of this correlation were significant ($r = .53$, $p < .01$). Thus, there was a strong, positive relation between using the notepad for problem solving and successful performance on these tasks.

Relation Between SEPS and Notepad Design

The previously described analysis supports our prediction that SEPS determines the effectiveness of note-taking for learning. However, it does not explain why the freeform notepad elicited superior performance to the structured and semistructured notepads. Recall that the freeform notepad allowed participants to take notes in whatever way they wanted, and thus it was extremely flexible. The structured notepad, on the other hand, was designed to guide participants toward a particular solution strategy (Identify Variable) by which we knew that the participants could solve the tasks. Given that our participants found these tasks quite difficult (as indicated by the low overall accuracy of the participants in the control group), one might well expect that providing a notepad that guided them to a "failsafe" solution strategy would indeed help them to perform better. Yet the structured notepad was the least effective of the notepad designs. What could account for this somewhat counterintuitive result?

To explore this question, we investigated the relation between the freeform and structured notepad conditions, SEPS uses of the notepad, and solution accuracy.[3] (For these analyses, we did not include the two cases in the freeform notepad condition for which a SEPS use was followed by a non-SEPS use.) First, using the coding scheme previously described, we ascertained the extent to which participants in the freeform and structured notepad conditions used the notepad for SEPS. In the freeform notepad condition, there were 14 participants who used the notepad on a total of 26 tasks. Of these 26 notepad uses, 18 (69%) were SEPS uses. Participants were correct on 14 of these 18 tasks (78%). In the structured notepad condition, there were 15 participants who used the notepad on a total of 27 tasks. There were 15 SEPS uses of the notepad, and participants were correct on 12 of these tasks (80%). Figure 16 summarizes these results. As Figure 16 shows, participants in the freeform notepad condition made slightly more use of SEPS strategies than those in the structured notepad condition; however, on the tasks for which participants did use SEPS, they were equally accurate.

Next, we examined the extent to which participants used the notepad for non-SEPS purposes. In the freeform notepad, participants merely copied information on 6 of the 26 tasks (23%). They were correct on 3 of those 6 tasks (50%). In the structured notepad condition, participants used non-SEPS strategies (Linear or Random Match) on 12 tasks (44%); however, they were correct on only 2 of those tasks (17%). Figure 17 summarizes the results for the non-SEPS uses of the notepad. Figure 17 clearly shows that when participants used the notepad for non-SEPS purposes, their performance was degraded. This was particularly the case in the structured notepad condition.

Together, these analyses show that when participants in the structured notepad condition used the notepad appropriately, they performed as well as those in the freeform notepad condition who used the notepad for SEPS. However, an almost equivalent number of participants who had access to this notepad did not appear to be able to take advantage of its structured property and, in these cases, performed extremely poorly on the tasks. Their performance was much worse than that of those participants in the freeform notepad group who used the notepad for copying, and their performance was even worse than that of participants

---

[3] We also analyzed SEPS use and performance among the semistructured notepad participants. This analysis was in keeping with the results for the freeform and structured notepad conditions; however, because the freeform and structured notepad conditions represent the extremes and for simplicity's sake, we report only the results from these two notepad conditions.

Figure 16. **SEPS use and accuracy for freeform and structured notepads.**

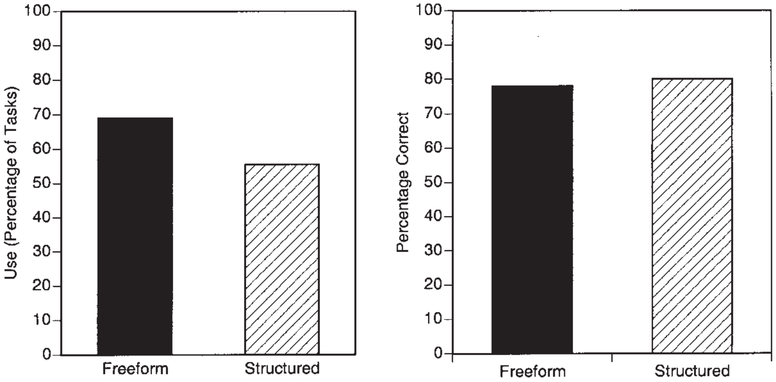
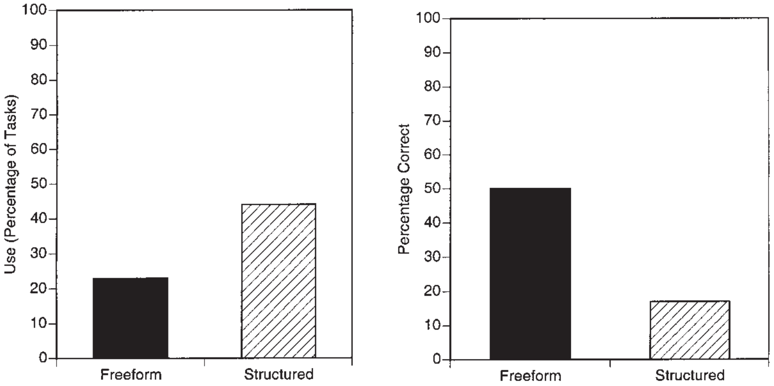


Figure 17. **Non-SEPS use and accuracy for freeform and structured notepads.**



in the control (no-notepad) condition. Their confusion about how to use the notepad seems to have significantly hurt their performance.

Participants who had access to the freeform notepad were given no guidance in how to use the notepad; they had to determine for themselves what kind of entry would be most beneficial. They used the notepad in a variety of ways, including several which mapped directly to strategies by which the problems could be solved, and they were successful on most of those tasks. When participants used the notepad merely to copy what was otherwise visible on the screen, they solved the problem only 50% of the time, about the same success rate as those in the control (no-notepad) condition. Failure to use the notepad for SEPS appears to have reduced their performance to that of the participants who did take notes at all.

4. GENERAL DISCUSSION AND CONCLUSION

The results of these studies suggest that taking notes can be a useful general strategy for learning. Experiment 1 presented correlational data suggesting that participants who took notes not only performed better than those who did not but also continued to perform better even after they stopped taking notes. Experiment 2 replicated the results of Experiment 1 and extended them by experimentally manipulating the participants' access to the notepad. In addition, Experiment 2 investigated the relation between the type of use of the notepad and problem-solving performance. Experiment 2 further investigated the effect on note-taking strategy of different designs of notepad.

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, the finding that the benefit of using the notepad carried over to tasks on which participants no longer used it suggests that taking notes actually can 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.

These results suggest that when engaged in problem solving, students may be helped by taking notes during the early stages of their problem-solving efforts—that is, while they are still trying to understand the task and the steps necessary to solve it. However, as important as note-taking may be during this phase of problem solving, it may not be as important after students understand what they are doing. Note-taking thus may be a scaffolding strategy, one that is helpful during the initial stages of problem solving but can be phased out as learners become more proficient at problem solving.

Finally, we found that although overall participants who took notes performed better than those who did not, using the notepad per se did not automatically lead to better performance and learning. We propose that the reason some uses of the notepad led to better performance whereas others did not is that the effective uses engaged the participants in SEPS. This was the case regardless of the type of notepad the participants used, and it appears to be the key to successful note-taking.

The results of our investigation of the relation between the design of notepad and the use participants made of it have implications for the design of note-taking facilities for learning and problem-solving environments. The freeform and the structured notepads represent two different approaches to supporting learning. By virtue of its unstructured nature, the freeform notepad requires that the user make decisions about what type of entry to make on the notepad. It therefore has significant potential to engage users in SEPS. However, this flexibility does not necessarily guide the user to use the notepad for SEPS. In several instances, when given this type of notepad, participants used it to copy what was already visible or even to make comments to the experimenter that were extraneous to the problem-solving task.

Although one might suppose that more scaffolding would lead to more use of the notepad for SEPS, this does not necessarily appear to be the case. The advantage of the structured notepad lay in its potential to serve as a highly effective tool that guided participants toward an appropriate solution strategy, and for some participants, this was the case. The disadvantage of this design was that when participants did not use the tool for SEPS, they used perceptually or randomly driven strategies that adversely affected their performance on the tasks.

One question arises from the results of this research: What is the extent to which our findings would generalize to different tasks, and especially to more complex tasks? The tasks used in our studies were scaled-down abstractions of scientific discovery tasks. However, our participants by no means found these tasks either simple or easy (i.e., in the no-notepad condition, participants in Experiment 2 solved less than 50% of the problems correctly). Nonetheless, we acknowledge that in most scientific discovery environments, students must manipulate many more variables than was required by our task, in addition to conducting many different kinds of observations and analyses. Whether a freeform notepad is optimal in more complex tasks remains an open question. In addition, many of the environments that use a notepad have many different tasks and problems for students to solve. Building a notepad for each problem type may become prohibitive for many complex systems.

Finally, we believe that the factor that determined the effectiveness of the notepad was the degree to which it supported SEPS, and that insofar as a notepad supports SEPS, it will enhance performance, regardless of the complexity of the task. In more complex problem-solving environments, it is possible that a different design of notepad might most effectively promote SEPS among learners and therefore would be more appropriate than a freeform notepad. How exactly such a notepad should be designed warrants further investigation.

A related issue is whether there was any value added by building the notepad into the computer interface. Given that in our studies, participants performed best when using the freeform notepad, would providing participants with a pencil and a blank sheet of paper have worked just as well? We believe that participants would have been less likely to make use of paper and pencil. In addition to needing to rearrange desk space to accommodate such a notepad (e.g., adjusting the position of the keyboard), they would have had to have taken their hands off the keyboard and moved their attention to the paper and pencil. The online notepad provided a convenient means of taking notes without diverting participants' attention from the task. Furthermore, after using a paper notepad, participants would be required to look in two places for information (the computer screen and the notepad) and to integrate information across these two locations. The online notepad kept their notes close to the area where they were working to solve the problems and had the additional advantage of automatically updating their notes.

Another large advantage of having the notepad built into the interface is the ability to track the students' progress on the task with the notepad, as was done in our studies. For example, a student who is having problems with a task may show some of his or her misconceptions on an online notepad, which then could be examined by a teacher or as part of an intelligent tutoring system (e.g., Koedinger, Suthers, & Forbus, 1999).

The results of this research suggest that it is useful to include a note-taking facility in the design of any learning or problem-solving environment but that this notepad should be designed with the potential not only to support SEPS but also to actively guide the learner to use such strategies. Insofar as the learner uses the notepad to engage in SEPS, performance and learning are likely to benefit. Using the notepad in other ways, however, is unlikely to enhance performance, and therefore the design of the notepad should prevent such non-SEPS uses.

The results suggest that having students take notes that engage them in self-explanation and problem solving is one means to improve their performance. Providing a note-taking facility that directs students toward SEPS in a problem-solving environment is an inexpensive and comparatively straightforward form of scaffolding that may have the further advantage of helping students learn general strategies by which problems can be solved.

NOTES

Background. An earlier version of Experiment 1 appeared *in The Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (Trickett & Trafton, 1999).

Acknowledgments. We thank Erik M. Altmann, Robert W. Holt, Irvin R. Katz, Audrey W. Lipps, Sheryl L. Miller, Lelyn D. Saner, Christian D. Schunn, and two anonymous reviewers for their comments. We give special thanks to Paula D. Raymond for her continuing comments on this research.

Support. This research was supported in part by Grant N0001400WX40002 from the Office of Naval Research to Greg Trafton and by a student fellowship from George Mason University to Susan Trickett.

Authors' Present Addresses. Greg Trafton, NRL, Code 5513, Washington, DC 20375–5337. E-mail: trafton@itd.nrl.navy.mil. Susan Trickett, Department of Psychology, George Mason University, Fairfax, VA 22030. E-mail: stricket@gmu.edu.

HCI Editorial Record. First manuscript received January 4, 2000. Accepted by Elliot Soloway. Final manuscript received October 27, 2000. — *Editor*

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4,* 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,* 91–95.

Barron, B. J., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., Bransford, J. D., & the Cognition and Technology Group at Vanderbilt. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences, 7,* 271–312.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4,* 253–278.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 849–874). Amsterdam: Elsevier Science B. V.

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

Kiewra, K. A. (1989). A review of note-taking: The encoding storage paradigm and beyond. *Educational Psychology Review, 1,* 147–172.

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

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

Koedinger, K., & Anderson, J. R. (1993). Reifying implicit planning in geometry: Guidelines for model-based intelligent tutoring system design. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 15–46). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8,* 30–43.

Koedinger, K. R., Suthers, D. D., & Forbus, K. D. (1999). Component-based construction of a science learning space. *International Journal of Artificial Intelligence in Education, 10,* 292–313.

Kozma, R. B., Russell, J., Jones, T., Marx, N., & Davis, J. (1996). The use of multiple, linked representations to facilitate science understanding. In S. Vosniadou, E. De Corte, R. Glaser, & H. Mandle (Eds.), *International perspectives on the design of technology-supported learning environments* (pp. 41–60). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Laidlaw, E. N., Skok, R. L., & McLaughlin, T. F. (1993). The effects of notetaking and self-questioning on quiz performance. *Science & Education, 77,* 75–82.

Linn, M. C. (in press). Designing the Knowledge Integration Environment: The partnership inquiry process. *International Journal of Science Education.*

Loh, B., Reiser, B. J., Radinsky, J., Edelson, D. C., Gomez, L. M., & Marshall, S. (in press). Developing reflective inquiry practices: A case study of software, teacher, and students. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from professional, instructional, and everyday science.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Milson, R., Lewis, M. W., & Anderson, J. R. (1990). The teacher's apprentice project: Building an algebra tutor. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 53–71). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

National Council of Teachers of Mathematics Commission on Standards for School Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Reiser, B. J., Kimberg, D. Y., Lovett, M. C., & Ranney, M. (1992). Knowledge representation and explanation in GIL, an intelligent tutor for programming. In J. H. Larkin & R. W. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (pp. 111–149). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Rinehart, S. D., & Thomas, K. F. (1993). Summarization ability and text recall by novice studiers. *Reading Research and Instruction, 32*(4), 24–32.

Roschelle, J. (1992). Learning by collaboration: Convergent conceptual change. *Journal of the Learning Sciences, 2,* 235–276.

Salzman, M. C., Dede, C. J., & Loftin, R. B. (1999). VR's frames of reference: A visualization technique for mastering abstract multidimensional information. *Proceedings of the CHI'99 Conference on Human Factors in Computing Systems*, 489–495. New York: ACM.

Sandoval, W. A., & Reiser, B. J. (1997, March). *Evolving explanations in high school biology.* Paper presented at the annual meeting of the American Educational Research Association (AERA'97), Chicago.

Sandoval, W. A., & Reiser, B. J. (1998, April). *Iterative design of a technology-supported biological inquiry curriculum.* Paper presented at the annual meeting of the American Educational Research Association (AERA'98), San Diego, CA.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28,* 859–882.

Schwarz, B., & Dreyfus, T. (1993). Measuring integration of information in multirepresentational software. *Interactive Learning Environments, 3,* 177–198.

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

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

Shute, V., & Glaser, R. (1991). An intelligent tutoring system for exploring principles of economics. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 333–366). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shute, V. J., Glaser, R., & Raghavan, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman & R. J. Sternberg (Eds.), *Learning and individual differences: Advances in theory and research. A series of books in psychology* (pp. 279–326). New York: Freeman.

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,* 360–370.

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

Slotta, J. D., & Linn, M. C. (in press). The knowledge integration environment: Helping students use the internet effectively. In M. J. Jacobson & R. Kozma (Eds.), *Learning the sciences of the 21st century.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Tabak, I., & Reiser, B. J. (1997, June). Domain-specific inquiry support: Permeating discussions with scientific conceptions. *Electronic Proceedings of "From Misconceptions to Constructed Understanding," The Fourth International Misconceptions Seminar,* Ithaca, NY.

Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Reiser, B. J. (1998, April). *Reflection as a vehicle toward local and global understanding.* Paper presented at the annual meeting of the American Educational Research Association (AERA'98), San Diego, CA.

Tabak, I., Smith, B. K., Sandoval, W. A., & Reiser, B. J. (1996). Combining general and domain-specific strategic support for biological inquiry. *Proceedings of the 3rd International ITS'96 Conference on Intelligent Tutoring Systems,* XXX–XXX. Montreal: PUBLISHER.

Trafton, J. G. (1994). *The contributions of studying examples and solving problems to skill acquisition.* Unpublished doctoral dissertation, Princeton University, Princeton, NJ.

Trafton, J. G., & Reiser, B. J. (1991). Providing natural representations to facilitate novices' understanding in a new domain: Forward and backward reasoning in programming. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society,* 923–927. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Trafton, J. G., & Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society,* XXX–XXX. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Trickett, S. B, & Trafton, J. G. (1999). Note-taking as a strategy for learning. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society,* 742–748. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Trickett, S. B., Trafton, J. G., & Raymond, P. D. (1998, August). Exploration in the experiment space: The relationship between systematicity and performance. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society,* 1067–1072. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

VanLehn, K., Niu, Z., Siler, S., & Gertner, A. (1998). Student modeling from conventional test data: A Bayesian approach without priors. *Proceedings of the 4th Intelligent Tutoring Systems ITS'98 Conference,* 434–443. Heidelberg, Germany: Springer-Verlag.

Van Meter, P., Yokoi, L., & Pressley, M. (1994). College students' theory of note-taking derived from their perceptions of note-taking. *Journal of Educational Psychology, 86,* 323–338.

Wason, P. C., & Johnson-Laird, P. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10,* 1–100.