Practice Effects and Learner Control on Acquisition, Outcome, and Efficiency

Valerie J. Shute
Armstrong Laboratory/HRTI
1880 Carswell Ave. Bldg 9020
Lackland AFB, TX 78236-5507
(210) 671-2734
vshute@colab.brooks.af.mil

Lisa A. Gawlick
Galaxy Scientific Corporation
1880 Carswell Ave. Bldg 9020
Lackland AFB, TX 78236-5507
(210) 671-2667
lgawlick@colab.brooks.af.mil

Abstract

This paper presents results from a study that attempted to replicate unexpected findings from a previous study (Shute & Gawlick, 1995) which investigated the effects of differential practice opportunities on skill acquisition, outcome, efficiency, and retention. These same variables were examined in a new study ($N = 380$), and the following results were replicated: (1) Learners receiving fewer practice opportunities completed the curriculum significantly faster than the other practice conditions, but at the expense of greater errors; and (2) Despite acquisition differences, all groups performed comparably on the outcome measure. This study also examines the effects of learner control (LC) on these same parameters. We included a condition where students chose their degree of practice, per problem set. Overall, this group completed the curriculum faster, and showed the highest outcome efficiencies, relative to the other conditions. Preliminary results from the retention part of this study ($n = 76$) continue to show an overall LC advantage, as well as a significant condition × gender interaction. That is, the LC condition is optimal for males, while the extended practice condition is best for females. We discuss the implications of these findings in relation to the design of efficacious instruction.

How does practice affect knowledge and skill acquisition, learning outcome, efficiency, and retention? On the one hand, there is a lot of support for the "practice makes perfect" position (e.g., Bryan & Harter, 1899; Schneider & Shiffrin, 1977). More recently, Anderson (1993) has provided compelling evidence for, and concluded that, *"Students achieve at higher levels if they solve more problems, whatever the regimen."* (p. 160). On the other hand, Schmidt and Bjork (1992) presented some interesting studies showing how, relative to a "standard" practice condition, acquisition conditions that slowed the rate of improvement, or decreased performance during practice, still yielded enhanced post-training performance. What is ultimately learned may therefore be obscured during the acquisition process, as relatively permanent effects become confounded with temporary performance effects that may disappear after the practice session is finished, or when the test conditions change.

The literature on learner control is even less definitive. Computerized learning environments can be characterized by the amount of learner control supported during the learning process. This dimension can be viewed as a continuum ranging from minimal (e.g., rote or didactic environments) to almost complete learner control (e.g., discovery environments). Two opposing perspectives address the issue of the best learning environment to build in intelligent instructional software. One approach is to develop an environment which provides the learner freedom to explore and learn (e.g., Collins & Brown, 1988; Shute, Glaser, & Raghavan, 1989). The other approach argues that it is more efficacious to develop directive learning environments (e.g., Corbett and Anderson, 1989; Sleeman, Kelly, Martinak, Ward, & Moore, 1989). Actually, this disparity may be resolved by, instead of looking for main effects of learning environment, additionally considering learner characteristics with the goal of identifying optimal learning environments for specific kinds of persons.

This paper reports the results from a large-scale study ($N = 380$) conducted to replicate previously-obtained (and unexpected) findings that also tested practice effects on skill acquisition, learning outcome, efficiency, and retention. In addition, we examine the role of learner control in relation to these same parameters. We report the results from phase I of the study that's been completed, and present preliminary results from a follow-on portion of the experiment where the same learners return, after 6 months, to see how much they remember, and if that differs due to original practice condition.

Our previous learning criterion task (Shute & Gawlick, 1995) was an intelligent tutoring system teaching flight engineering knowledge and skills, divided into two main curriculum sections. Each section had two alternative conditions, differing only in the number of practice opportunities across problem sets: "Abbreviated" (A) and "Extended" (E). Thus, there were four practice conditions: AA, AE, EA, EE. Despite differences in acquisition (i.e., learners in the abbreviated conditions made more errors during problem solution compared to the other groups receiving more practice opportunities), groups performed the same across *all* learning outcome measures (surprise #1). We speculated that practice effects, while not readily apparent, may show up after some period of time had elapsed. In fact, the second experiment showed evidence for practice effects on long-term retention (i.e., after more than *two years*), but not in the predicted direction (surprise #2). That is, learners in the mixed conditions (switched 3/5 of the way through the curriculum from one practice condition to another) showed significantly greater retention compared to those assigned to either of the two homogeneous conditions.

In the current study, we use the same four treatment conditions as in the previous study, but have employed a completely different instructional environment (i.e., Stat Lady, teaching introductory statistics) to test the generalizability of the previous findings in a different domain. Furthermore, we include a fifth treatment condition, Learner Control (LC), which allows learners to select the number of problems to solve, per problem set, rather than solving a fixed number of problems. By including this new treatment condition, we can test the effects on these same learning parameters when learners are in control of their practice opportunities. Do individuals, in general, have the necessary metacognitive skills to know when additional help is needed, or when they've had enough practice? Are there individual differences in terms of who benefits most by this condition?

Hypotheses

Skill Acquisition. Based on previous findings, and given fewer practice opportunities in which to apply newly-developing knowledge and skills, we expected learners assigned to the more limited practice conditions to exhibit more errors during learning compared to those learning from the more extended conditions. We further expected learners in the LC condition to perform about average during skill acquisition, making a moderate number of errors compared to the other conditions. This was based on the belief that these learners would elect to solve a large range of problems due to individual differences in general aptitude, metacognitive skills, and personality traits. The result was expected to balance out at a middle level of performance.

Learning Outcome. We predicted no differences on the posttest measure among groups, given findings from the previous study. However, if there *were* any differences, we expected learners in the most extended conditions to perform better on the outcome measures compared to learners in the abbreviated conditions given they would have had significantly more practice opportunities (Anderson, 1993). With regard to the LC condition, we speculated that these individuals would show an intermediate level of outcome performance given greater variability in the number of problems they chose to solve.

Learning Efficiency. The time taken to complete the tutor should be a direct function of practice condition. Thus, learners in the most abbreviated conditions would take the least amount of time to complete the curriculum given fewer problems to solve, and learners in the most extended conditions would take the most amount of time. Learners in the LC condition were expected to take an intermediate amount of time as we believed that learners are often not cognizant of their cognitive strengths and weaknesses, nor are many of them sufficiently motivated to continue practicing until a skill is mastered.

Retention. On the basis of our earlier findings (Shute & Gawlick, 1995), we hypothesized that learners in the mixed practice conditions would show greater retention of the material compared to learners in the homogeneous conditions following a 6-month lag between original and retention testing. We also hypothesized that learners originally assigned to the LC condition would show average, to above-average levels of retention based on a fairly typical finding in the learner-control literature which suggests that increased control over one's environment renders the learning experience more enjoyable, particularly for high-ability learners (e.g., Hannafin & Sullivan, 1996; Shute & Gawlick-Grendell, 1994; Swanson, 1990). Finally, in addition to testing for main effects of condition on retention, we were interested in examining the role of gender; specifically in terms of a possible interaction with condition. While we did not expect to see a main effect of gender, we did posit a gender × condition interaction whereby males were expected to show greater retention having learned in the LC condition (compared to the other conditions), and females to show better retention having learned from more extended practice conditions. This hypothesis was motivated by Shute & Gluck (in press) who reported that males showed significantly more independent/exploratory behaviors than females when learning from an on-line instructional system, and this particular tendency would be well-suited to the LC condition, possibly resulting in increased retention.

Method

Participants

A total of 380 individuals participated in this experiment, obtained from local temporary employment agencies. The age range of the sample was between 18-30 years (Mean = 22), and all had a high school diploma or equivalent. Overall, 66% of the sample was male, and no one had any prior exposure to statistics courses. Participants were paid for taking part in the study and informed that they needed to return in 6 months for phase 2--retention testing. To motivate their return, we offer a monetary bonus. Currently, we have collected data from a total of 76 individuals who have returned for the second part of the study.

Materials

The first module of the Stat Lady Descriptive Statistics series (DS-1, Shute & Gluck, 1994) was used as the complex learning task in the experiments described in this paper (for more on this module, see Shute, 1995). The curriculum was decomposed via a cognitive task analysis into curriculum elements (CEs), representing low-level bits of knowledge and skill (e.g., identify the symbol for summation, sum all frequencies in a given sample). In this study, participants received 77 CEs, arranged from simple to more complex concepts and skills, and spread across five main problem sets or topics: (a) frequency distributions, (b) proportions and percentages, (c) grouped frequency distributions, (d) cumulative frequency distributions, and (e) plotting. Each CE (or small group of related CEs) was instructed by Stat Lady, then individuals were assessed for CE mastery on the basis of their problem solving performances. In this study, the number of problems that learners solved (per problem set) was solely a function of assigned condition. All learners had to solve between 1 to 4 CE-related problems before moving on to the next problem set. If a learner gave an incorrect answer or solution, Stat Lady intervened with progressively more specific feedback related to the particular error. Learners were allowed up to three errors before Stat Lady provided the correct answer.

Because each CE was directly mapped to a specific question/problem (note: some CEs had several associated questions), participants in the extended condition received three more questions per CE than learners in the abbreviated condition (maintaining a 4:1 ratio between extended and abbreviated practice opportunities). At the end of the tutor, the system computed each learner's average number of questions and errors, per CE. The "questions" variable was constant for learners in the fixed practice conditions (but varied for the LC condition) while the "errors" variable differed for all learners, reflecting degree of problem-solving difficulty. Stat Lady's three-level feedback thus allowed learners to make between 0 to 3 errors, per question.

To assess learning outcome, duplicate items were created to assess knowledge/skill related to each of the 77 CEs. This resulted in two parallel forms of a test (A and B) that were administered on-line, before and after the tutor. For more details on the specifics of these tests, see Shute (1995).

Design and Procedure

In the previous study, participants were either switched to a new practice condition (e.g., A→E), or remained in the same one (e.g., E→E) about 3/5 of the way through the tutor. Similarly, in this study, learners (not in the LC condition) were either switched to a new condition or remained in the same one, after the 3rd (of 5) problem sets. Thus, there were a total of 5 practice conditions: (a) AA (n=86), (b) AE (n=60), (c) EA (n=58), (d) EE (n=88), and (e) LC (n=88).

Participants were tested in groups of about 20, and randomly assigned to a condition. Given the two parts of the tutor and the 4:1 ratio described above, the total number of problems presented, per condition, were: AA (5), AE (11), EA (14), EE (20), and LC (variable, between 5 - 20).

On-line demographic questionnaires and pretests were administered to all participants. After completing both, they proceeded to learn from the tutor which took, on average, about 5 hr to complete. Finally, all participants were administered an on-line posttest assessing the full range of knowledge and skills acquired from the tutor.

Participants in the first phase of this study were asked to return 6 months after learning from Stat Lady to take part in the follow-up portion (phase 2) of the study. Currently, 20% of the original sample has returned (n=76). The average lag between original and retention testing = 26.3 weeks (SD = 2.7 wk). The distribution of the returning participants, by condition, is: AA (n=13), AE (n=12), EA (n=16), EE (n=12), and LC (n=23).

Testing for the retention part of the study is being conducted in small groups of about 5 persons, over one day. Prior to taking the first retention test (consisting of items which are isomorphic to those used in phase 1), test administrators brief each group on the importance of trying to remember as much as they can from their original session. After the first test has been completed, participants are given a 30-minute break, followed by the second retention test. At the conclusion of the second test, all returning participants are administered an on-line battery of cognitive ability tests assessing working memory capacity, information processing speed, inductive reasoning skill, and fact learning ability, in the quantitative domain. This battery requires, on average, about one hour to complete.

Results

Prior to making comparisons between practice conditions, we needed to insure that learners within each condition were demographically comparable. Several one-way ANOVAs were computed on age, gender, number of years of education, and computer experience, by condition. None of these variables showed significant differences across the five practice conditions.

Skill Acquisition

Does practice condition affect acquisition accuracy? We examined this issue first by comparing the *number of errors* made during learning, averaged across all CEs. As mentioned, this value could range from 0 to 3 errors, per CE. Significant differences were found: $F(4, 368) = 7.46$, $p < .001$. By condition, the order of average errors was: LC (1.67) < AA (1.70) < AE (2.41) < EA (2.47) < EE (2.96). However, learners in the Extended condition received four times as many questions per CE compared to learners in the Abbreviated condition, so their "error" values should be considered in relation to the number of problems they solved (note: the average number of questions that LC learners chose to solve fell midway between AA and AE conditions). Thus, to test for differences in *acquisition accuracy*, we computed a new variable--the number of errors divided by the number of questions, averaged across CEs. For this index, values close to 1.0 denote average performance; values *less than* 1.0 denote more accurate performance (fewer mistakes relative to the number of questions) and values *greater than* 1.0 denote more inaccurate performance (more errors relative to questions received).

For our sample, this value ranged from 0.40 to 2.09 and was significantly different among conditions: $F(4, 362) = 20.45$, $p < .01$. The order of this variable by condition was: EE < EA < AE < LC < AA. Thus, as with the previous study, the EE learners' acquisition accuracy was the highest among conditions--they made fewer mistakes relative to their greater number of questions. Learners in the AA condition showed the lowest acquisition accuracy--they tended to commit more mistakes on relatively fewer questions. Learners in the LC condition showed accuracy indices that were about midway between the AA and the EE conditions. See Figure 1.

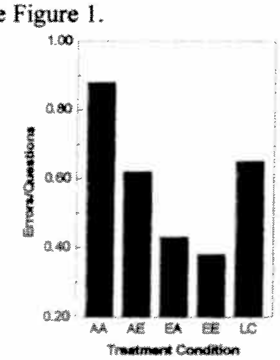

Figure 1: Acquisition accuracy across treatment conditions.

Learning Outcome

We first examined the pretest data to insure the five groups were comparable in prior knowledge and skills related to statistics. We computed an ANOVA on pretest scores, by condition. Although there were no significant differences among conditions, $F(4, 375) = 2.32$, $p = .06$, the F-value was sufficiently large to justify controlling for pretest data in subsequent analyses. Specifically, the EE group (by chance) began with the highest pretest Mean (greatest incoming knowledge), the AA with the lowest, and the LC participants, in between.

An ANOVA was computed on the posttest data (Means adjusted for pretest score) by condition, and there were no significant differences: $F(4, 375) = 2.94$, $p = .06$. However, as predicted, the order of posttest scores was: AA (68.3) < AE (71.4) < LC (71.9) < EA (74.5) ≈ EE (74.6).

Learning Time

We decomposed the total tutor-time variable into two parts--instruction and problem-solving time, representing two distinct parts of the Stat Lady program. Instruction time should vary in relation to one's facility in acquiring and understanding the new material, while problem-solving time should vary in relation to condition.

Table 1: Instruction, problem-solving, and overall tutor time (hrs) by condition.

Condition	Instruction	Prob-Solving	Total
AA ($n = 86$)	2.05	1.98	***4.03***
AE ($n = 60$)	1.88	2.96	***4.84***
EA ($n = 58$)	1.67	3.16	***4.85***
EE ($n = 88$)	1.52	3.96	***5.47***
LC ($n = 88$)	1.78	2.13	***3.91***
F (4, 375)	8.78	40.38	***12.74***
p	< 0.001	< 0.001	***< 0.001***

Notes: A = Abbreviated, E = Extended, and LC = Learner Control practice conditions.

Three ANOVAs were computed on instruction time, problem-solving time, and total time required to complete the tutor. All three variables showed significant differences due to condition. The order of total time by condition was unexpected: LC < AA < AE = EA < EE. Contrary to our hypothesis, the LC learners were fastest of all (see Table 1).

The final variable that we examined combined outcome score (i.e., adjusted posttest data) and tutor-completion time to yield an *outcome-efficiency* index (i.e., posttest score divided by time on tutor). The interpretation of this variable is that larger values reflect greater efficiency (i.e., higher learning outcome scores relative to time spent on the tutor). Lower values indicate less efficient learning.

We computed an ANOVA on this ratio by condition and the results were significant: $F(4, 375) = 6.00$, $p < .001$. The ordering of this index, by condition, was: EE < AE ≈ EA < AA < LC. As can be seen in Figure 2, LC learners showed superior learning efficiency relative to the other conditions.

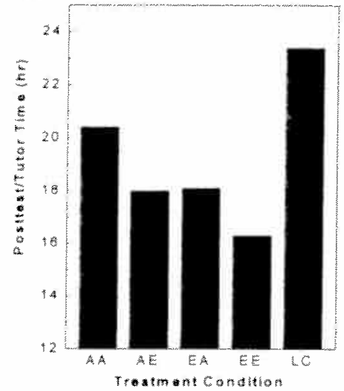

Figure 2: Learning outcome efficiency data by condition.

Retention (Preliminary Results)

Currently, we have data from $n = 76$ of the original $N = 380$ participants for the retention part of the study (i.e., phase 2). This phase will be completed June 1996. The question here is how the practice conditions, in general, affect retention of this material. Because this represents an incomplete study, the following should be viewed as preliminary analyses and tentative conclusions.

Prior to making comparisons among conditions on retention, we needed to insure that the subset of returning learners were comparable to the original sample (phase 1 data overall, and per practice condition). We computed one-way ANOVAs on demographic measures (age, gender, education, computer experience), by phase. None of these measures were significantly different. We also compared returning to original participants' data on phase 1 posttest scores (adjusted for pretest). Scores from the returning sample did not differ significantly from the original sample on this measure, $t(454) = -0.11$, $p = .91$.

Next, we computed a factor analysis (principal components analysis) on the cognitive ability test data (percent correct scores). This resulted in the extraction of a single factor: general aptitude. The percentage of variance accounted for by this factor was 64.0%, with $M = 0$, and $SD = 1$. Factor scores were saved for each person and used as a covariate in subsequent analyses.

We then combined data from individuals originally learning from the AE and EA conditions because: (a) their acquisition, outcome, and efficiency data from phase 1 were not significantly different to warrant their separation, (b) this increases the power of the upcoming analyses, and (c) this same procedure was followed in the original Shute & Gawlick (1995) study. Furthermore, we combined the two retention test scores into an average retention score.

To test our hypotheses concerning condition and gender effects on retention, we computed an ANOVA with *retention* as the within-subjects variable (i.e., the adjusted posttest scores from phase 1 and average retention scores from phase 2 as the repeated measures). *Condition* (AA, AE/EA, EE, LC) and *gender* (male, female) were the between-subjects variables. We included the *aptitude* factor score as a covariate in the equation to control for any differences in aptitude that may mediate any obtained main effects or interactions. Results from the ANOVA showed no main effect on retention due to original practice *condition* ($F < 1$), no main effect of *gender* ($F < 1$), but a significant *condition* × *gender* interaction: $F(3, 67) = 3.79$, $p = .01$. This interaction is depicted in Figure 3.

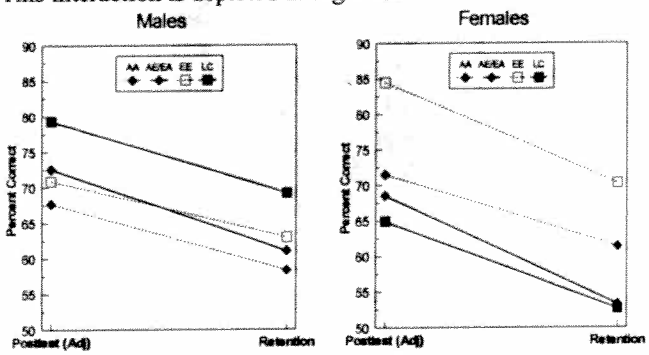

Figure 3. Condition by gender interaction on retention

Finally, we created a *retention-efficiency* index--retention score (with adjusted posttest score from phase 1 partialled out and the retention score residuals saved for each person) divided by original learning time. Again, higher numbers mean greater retention relative to acquisition time. We computed an ANOVA on this index by condition and the results were significant: $F(3, 72) = 3.54$, $p = .02$. By condition, the ordered indices were: EE (13.4), AE/EA (14.2), AA (18.7), and LC (24.6). Comparing just the relative indices related to the LC and EE conditions, the effect size = 1.6, with a strong LC advantage over the EE condition.

Discussion

Is it really the case that more practice opportunities yield better achievement, regardless of regimen; or is the relationship more complicated? The first purpose of this study was to replicate rather unexpected findings from our original study (Shute & Gawlick, 1995) that tested this query. Disregarding the learner control manipulation for a moment, we specifically replicated the following: (a) Reduced practice opportunities result in worse accuracy acquisition, but (b) despite these acquisition differences, outcome performances across all conditions are equivalent (even in this different domain), and (c) learners in the abbreviated condition(s) complete the tutor significantly faster than learners in the more extended conditions. Moreover, when we view tutor-time data separated into its component parts (instruction and problem-solving time), we see, predictably, that problem-solving time increases as a function of practice condition (AA < AE < EA < EE). However, instruction time shows a *reversal* of this ordering: EE < EA < AE < AA. This suggests that abbreviated learners may have been attempting to compensate for their sparse practice environments by spending relatively more time reviewing the instructional sections of the tutor.

A second goal of phase 1 of this study was to examine the effects of learner control on these same parameters (i.e., acquisition, outcome, and efficiency). While the LC learners did show an intermediate level of acquisition accuracy, surprisingly, they completed the tutor faster than any other condition. And when we computed and tested an outcome-efficiency index by condition, results showed the LC learners greatly surpassed the other groups (see Figure 2).

In the second phase of this study, we are (a) attempting to replicate findings from a previous retention study (Shute & Gawlick, 1995), where greater retention of flight engineering knowledge and skills were exhibited by those who originally learned from mixed-practice conditions (AE and EA), and (b) examining the effects of practice condition and gender on retention.

Following six months between original instruction (phase 1) and retention testing (phase 2) in the present study, using a different domain, we found no significant main effect of original practice condition on retention, thus we failed to replicate the finding of the mixed-practice condition's advantage on retention (although the data are still incomplete). However, we have found a very interesting condition by gender interaction. Figure 3 clearly shows that for males, learning from the LC condition represents the superior learning environment--their outcome and retention scores are much higher compared to males learning in the other practice conditions. The female data reveal a very different story--females learn better and remember more from the consistently extended condition (EE); their poorest performances are associated with the LC condition. This interaction is even more compelling in that it appears even with aptitude being controlled in the equation.

What would account for the obtained gender by condition interaction in relation to differential retention, particularly for the LC condition? First, aptitude-treatment interaction (ATI) research has shown that certain learner characteristics are better suited to specific kinds of environments to achieve optimal outcome performance (see Shute, Glaser, & Raghavan, 1989; Tobias, 1989, 1994). Second, Shute (1993) reported that individuals demonstrating greater exploratory behaviors perform better in more open learning environments (similar to the LC condition, and contrasting with more didactic ones) while the converse was found for less-exploratory individuals. Third, a different study (Shute & Gluck, in press) further examined exploratory behaviors in terms of optional, on-line tool usage, and reported gender effects related to tool use. That is, males tended to more spontaneously employ the on-line tools compared to females, and there was a main effect of tool use on learning outcome (i.e., more was better, overall). Finally, exploratory and independent kinds of behaviors have been linked to endogenous testosterone level, and males have significantly more testosterone than females (e.g., Broverman, Klaiber, Kobayashi, & Vogel, 1968; Kimura, 1992; Newcombe, 1982). Testosterone affects brain functions in a manner similar to an adrenergic stimulant--exerting an influence on precisely those traits that are best suited to a learning environment offering more learner control.

Obviously, more research is needed to test all of these relationships. We are currently completing a series of six studies that examine gender effects across a range of learning environments and domains, assessing testosterone levels, and relating these variables to performance. Many new gender by treatment interactions are emerging that have direct implications for adaptive instructional design.

In conclusion, participants in the LC condition complete the tutor much faster than the other fixed-practice conditions, yet perform no differently on the outcome measure compared to those having the most extensive practice opportunities. Moreover, we continue to see advantages for the LC condition after six months, illustrated by this group's large retention-efficiency index relative to the other conditions. The only downside is that this condition appears to be better for males than females, at least with regard to retention. Overall, these findings suggest that the design of automated instructional systems may be enhanced, and learning efficiency improved, by providing greater student control during learning. Furthermore, females may benefit by receiving pre-training that specifically focuses on ways to improve self-monitoring skills. In fact, we have data that suggests that, within the LC condition, males tend to adjust their requests for problems in relation to perceived need (i.e., asking for additional questions in relation to more difficult CEs, and fewer questions for less problematic CEs). On the other hand, our data show that females, in general, tend to ask for fewer problems. We plan to explore these issues in greater detail once all of the retention data have been collected.

Acknowledgments

We'd like to thank Kevin Gluck, Barry Goettl, Nancy Lefort, Jason Miller, Wayne Crone, Linda Robertson-Schüle, Rickard Robbins, Cathy Gomez and Shirley Snooks for their different (and excellent) contributions to this study.

The research reported in this paper was conducted as part of the Armstrong Laboratory, TRAIN Project, Brooks AFB, TX. Funds for this research were provided by the Air Force Office of Scientific Research. The opinions expressed in this article are those of the authors and do not necessarily reflect those of the Air Force. Correspondence concerning this paper should be addressed to the first author.

References

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Broverman, D. M,. Klaiber, E. L., Kobayashi, Y., & Vogel, W. (1968). Roles of activation and inhibition in sex differences in cognitive abilities. *Psychological Review*, 75, 23-50.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, 6, 345-375.

Collins, A. & Brown, J. S. (1988). The computer as a tool for learning through reflection. In H. Mandl and A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 1-18). New York, NY: Springer-Verlag.

Corbett, A. T. & Anderson, J. R. (1989). Feedback timing and student control in the Lisp intelligent tutoring system. In D. Bierman, J. Brueker, & J. Sandberg (Eds.), *Proceedings of the 4th International Conference on Artificial Intelligence and Education* (pp. 64-72). Springfield, VA: IOS.

Hannafin, R. D., & Sullivan, H. J. (1996). Preferences and learner control over amount of instruction. *Journal of Educational Psychology*, 99(1), 162-173.

Kimura, D. (1992). Sex differences in the Brain. *Scientific American*, 119-125.

Newcombe, N. (1982). Sex-related differences in spatial ability. In M. Potegal (Ed.), *Spatial ability: Development and physiological foundations* (pp. 223-250). New York, NY: Academic Press.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207-217.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. Detection, search, and attention. *Psychological Review*, 84, 1-66.

Shute, V. J. (1993). A comparison of learning environments: All that glitters... In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 47-74). Hillsdale, NJ: Erlbaum.

Shute, V. J. (1995). SMART: Student Modeling Approach for Responsive Tutoring. *User Modeling and User-Adapted Interaction*, 5, 1-44.

Shute, V. J. & Gawlick, L. A. (1995). Practice effects on skill acquisition, learning outcome, and retention. *Human Factors*, 37(4), 781-803.

Shute, V. J. & Gawlick-Grendell, L. A. (1994). What does the computer contribute to learning? *Computers and Education: An International Journal*, 23(3), 177-186.

Shute, V. J., Glaser, R. & Raghavan, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 279-326). New York, NY: W. H. Freeman.

Shute, V. J. & Gluck, K. A. (in press). Patterns of Exploratory Behavior: Causes and Effects. *The Journal of the Learning Sciences.*

Shute, V. J., & Gluck, K. A. (1994). *Stat Lady: Descriptive Statistics Module 1.* [Unpublished computer program]. Brooks Air Force Base, TX: Armstrong Laboratory.

Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science*, 13(4), 551-568.

Swanson, J. H. (1990, April). The effectiveness of tutorial strategies: An experimental evaluation. *Paper presented the American Educational Research Association*, Boston, MA.

Tobias, S. (1989). Another look at research on the adaptation of instruction to student characteristics. *Educational Psychologist*, 24(3), 213-227.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, 64(1), 37-54.