Effects of Practice and Learner Control on Short- and Long-Term Gain and Efficiency

Valerie J. Shute, U.S. Air Force Research Laboratory, Brooks Air Force Base, Texas, **Lisa A. Gawlick**, Galaxy Scientific Corporation, Lackland Air Force Base, Texas, and **Kevin A. Gluck**, Carnegie-Mellon University, Pittsburgh, Pennsylvania

This study investigated the effects of practice opportunities and learner control on short- and long-term learning from a computer-based introductory statistics curriculum. In all, 380 participants were assigned to one of five conditions. The first four conditions differed in terms of the number of problems to solve per problem set. The fifth condition allowed learners to choose the amount of practice. A subset (n = 120) of the original participants returned for testing following a six-month interval. Overall, the fixed-practice conditions showed learning gains that varied in relation to the amount of practice (i.e., more was better). The data from the learner-control condition was unexpected, showing learning gains comparable with the most extended practice condition yet the fastest tutor-completion times. We discuss implications of these findings in relation to the design of efficacious instruction. Actual or potential applications of this research include the modification of computer-based instruction that can enhance individuals' learning efficiency and outcome scores.

INTRODUCTION

Two independent research issues motivated the work reported here. First, we were interested in trying to replicate findings from an earlier study (Shute & Gawlick, 1995) that examined the effects of practice condition on performance, using a different domain in this study. Second, we wanted to compare learning results between individuals assigned to either a computer-controlled or learner-controlled condition during computer-based instruction. In order to establish a common theoretical and lexical foundation with the reader, we will summarize both of these motivations in more detail before describing the study.

Practice Effects

There is considerable support, both anecdotal and empirical, for the idea that "practice makes perfect," or, in its less extreme form, that "practice makes better" (Bryan & Harter, 1899; Newell & Rosenbloom, 1981; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The historical foundation for contemporary research on this topic was established by Thorndike's (1898) investigations at the turn of the century into the effects of practice with feedback. More recently, Anderson (1993) provided compelling evidence for the conclusion that "students achieve at higher levels if they solve more problems, whatever the regimen" (p. 160). The common conclusion across all of this work, oversimplified, is that the more often people perform a task, the more accurate and faster they become.

Agreement on the nature of this relationship between practice and performance is so universal that contradictory results often generate incredulity among the research community. There are, however, exceptions. Schmidt and Bjork (1992), for instance, presented an

Requests for reprints should be sent to Valerie J. Shute, Air Force Research Laboratory/HEJT, 1880 Carswell Ave. Bldg. 9020, Lackland AFB, TX 78236-5507; vshute@colab.brooks.af.mil. **HUMAN FACTORS**, Vol. 40, No. 2, June 1998, pp. 296–310.

excellent review of 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 posttraining performance (in both motor- and verbal-learning paradigms). In addition, the first two authors of this paper completed a study on practice effects that reported anomalous findings.

Shute and Gawlick (1995) used a computer-based instructional system for teaching flight engineering knowledge and skills to examine the effects of variable levels of practice on a number of learning measures, including immediate outcome and long-term retention. This earlier study consisted of four fixed-practice conditions, with a 4:1 ratio of practice opportunities between the most extended and abbreviated conditions (i.e., learners in the extended condition were required to solve four times as many problems as those in the abbreviated condition). The first unexpected result was that learners in all conditions performed equally well on the posttest.

It was hypothesized that practice effects would surface after some time had elapsed. To test that notion, a subset of learners was retested approximately two years following initial instruction. This produced the second unexpected result: Learners in the mixed-practice conditions, switched three-fifths of the way through the curriculum from abbreviated practice to extended practice (or vice versa), showed significantly greater retention compared with those assigned to either of the two homogeneous conditions (i.e., always abbreviated, always extended). Both of these findings differ from what one would expect, given that more practice is generally better than less.

In the current study, we used four treatment conditions that maintained a similar 4:1 ratio between extended- and abbreviated-practice groups, but we employed a different instructional environment to test the generalizability of the previous findings. That is, we wanted to examine the same practice schedules in a new domain and determine whether the effects of amount of practice yielded the same results. We also included a fifth condition, learner control. This allowed learners to select the number of problems to solve per problem set as they moved through the computer-based instruction, rather than having to solve a fixed number of problems presented by the computer. We next address the issue of learner control in computer-based learning environments.

Learner Control

The second research issue motivating the work described here involves the effect on learning when individuals are given control of their practice opportunities. Specifically, what effect, if any, does learner control (LC) have on short- and long-term retention and learning efficiency? The literature on the benefits of learner- versus computer-controlled (CC) conditions is enormously complex, and the results are divided. One difficulty in interpreting the mixture of conclusions in this literature is that various researchers define learner control differently and instantiate it in their instructional programs in different ways.

In the context of this paper, we define learner control as the ability to self-determine how many practice problems to complete before moving on to a new topic in the curriculum. Note that the emphasis is on control over practice problems, as opposed to, for instance, control over which instructional materials to see. The current state of confusion regarding issues of learner control should become apparent as we take a brief look at results across three dependent measures: learning outcome, learning time, and learning efficiency.

Learning outcome. Both Gray (1987) and Lee and Lee (1991) compared differences in learning as a function of whether the learner or the computer controlled the number of practice opportunities. These studies showed significantly better performance by the LC participants. A standard explanation for this result is that because each individual is best suited to assess his or her own learning needs and interests, the most efficacious approach is to give students lots of latitude in which to tailor their own learning experience (Merrill, 1975, 1980). Further support for the pro-LC interpretation comes from Hannafin (1984), who stated that a greater effect on learning and memory is seen when learners have control over their learning experience.

Other studies have reported the opposite finding – that learners show significantly poorer achievement when given control of various instructional elements (e.g., Belland, Taylor, Canelos, Dwyer, & Baker, 1985; Dalton, 1990; Tennyson, Park, & Christensen, 1985; Tennyson, Welsh, Christensen, & Hajovy, 1985). A typical explanation for the CC superiority is that learners are unskilled at determining successful strategies for themselves and the environment in which they are learning. Some believe that learners lack the necessary self-regulation skills for effective monitoring of their own knowledge and skill acquisition (Kinzie, 1990; Steinberg, 1989) or that they just make inappropriate decisions, such as ending their session too soon (see Carrier, 1984).

With these conflicting results established, we can add that many LC-versus-CC studies have reported no differences in learning outcome. For instance, Murphy and Davidson (1991) compared the effects of three treatment conditions (learner control, adaptive control, and learner advisement) on concept acquisition by nurses studying different types of shock. They found no significant differences among groups in terms of immediate recall, intermediate retention (two weeks), or long-term retention (six to eight weeks). This null finding has been interpreted as support for LC environments because the time and cost of programming lesson paths (curricula) for different individuals would be eliminated in an LC environment. Thus, if there is no difference between these two environments with regard to learning outcome, it would be harmless to let students handle their own lesson branching (Williams, 1996). Another sense in which time becomes a factor in reaching conclusions about instructional implications concerns learning time, which generally defined as the time spent learning from the tutor.

Learning time. The literature on learning times and efficiency is also ambiguous. In two studies comparing the learning times of pairs of students in LC and CC conditions (Dalton, 1990; MacGregor, 1988), dyads in the LC condition spent more time in the program than did those in the CC condition. This was attributed to a greater degree of socializing between pairs in the LC condition. The authors of those studies did not, however, conclude that students in each condition were more or less cognitively engaged in the learning task. A similar result regarding learning time was reported by Shyu and Brown (1992), who found no outcome differences between LC and CC conditions but did find that more time was spent by LC than CC participants. They attributed the extra time needed by the LC learners to their additional requirement of learning the control features of the interface before learning the curriculum.

Some studies have not found any differences between LC and CC conditions in time spent learning (Hurlock, Lahey, & McCann, 1974; Kinzie & Sullivan, 1989; Lahey, Hurlock, & McCann, 1973). However, the majority of the studies we reviewed have reported that LC students take less time to complete a curriculum than do CC students (e.g., Johansen & Tennyson, 1983; Murphy & Davidson, 1991; Tennyson, 1980; Tennyson & Buttrey, 1980). For instance, the Murphy and Davidson study reported no differences in outcome but did report faster completion times for LC than for CC participants. Other studies have similarly reported that LC learners demonstrate faster completion times but that learners in the CC condition end up with greater learning outcomes (e.g., Johansen & Tennyson, 1983; Rivers, 1972; Tennyson & Buttrey, 1980; Tennyson, Tennyson, & Rothen, 1980). Such results highlight the independence of these two dependent variables and serve as a reminder that another valid way to compare LC and CC conditions is in terms of efficiency indices (e.g., achievement per time spent).

Learning efficiency. Here again, the findings are mixed. In Goetzfried & Hannafin's (1985) study, for example, the LC condition was found to be less efficient than the CC condition because learners progressed more slowly through the instruction; however, Relan (1991) found partial LC to be the most efficient method of study. One major problem with interpreting LC in relation to instruction time is that LC learners might elect to skip important material, which would affect their performance (Lepper, 1985; Williams, 1996). Thus, the possible interactions among instructional control, time on task, and amount of instruction can obscure the interpretation of results.

Some researchers suggest allowing control over a few selected variables (e.g., context, sequencing, and presentation style) for learners less experienced in knowing what is best for themselves (Higginbotham-Wheat, 1988; Ross & Morrison, 1989). In our study, we eliminated this interaction by holding instructional material constant across all conditions (i.e., all learners view the same instruction). Furthermore, learners in all conditions controlled optional features in the environment, such as accessing the on-line dictionary and formula bank. The primary difference is that learners in the LC condition controlled the number of practice opportunities per problem set, whereas that variable was fixed for those in the CC conditions.

Results are reported from two phases of the study. The first phase involved learners acquiring new knowledge and skills from a computer-based statistics tutor called Stat Lady (Shute & Gluck, 1994). Participants were placed in one of the five different treatment conditions for this first phase and assessed in terms of short-term retention. The second part of the study involved a subset of the same participants, who returned after six months for long-term retention assessment.

METHOD

Participants

The 380 participants (72% men, 28% women), who were obtained from local temporary employment agencies, ranged in age from 18 to 30 years (mean = 22). All had a high school diploma or equivalency degree, and none had more than minimal prior exposure to statistics courses or training. Participants were paid $6/h for taking part in the study and were informed that they needed to return in six months for Phase 2 (retention testing). Participants were told that all who returned for retention testing would receive a $50 bonus in addition to their hourly wage.

Materials

The tutor. For the curriculum (i.e., criterion task) in this study, we used the first module of the Stat Lady Descriptive Statistics series (DS-1; Shute & Gluck, 1994). This module provides the basis for later modules that instruct measures of central tendency and variability. (For more detail on this module, see Shute, 1995.) The DS-1 curriculum consists of 77 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) derived from a cognitive task analysis of the domain "Data Organization and Plotting" (the subtitle of the tutor module). This analysis was performed in order to make explicit the instructional goals of the tutor (i.e., the 77 CEs), a prerequisite to the design of instruction, assessment, and test materials.

Two subject-matter experts independently categorized each CE into one of three knowledge-outcome types: symbolic knowledge, conceptual knowledge, and procedural skill. The agreement between the experts was high (r_{xy} = .95), most likely attributable to the concrete operational definitions and examples summarized and related to each outcome type. The few instances of categorization dissension were easily resolved by discussion. Within each of these three categories, the CEs were subsequently arranged in inheritance hierarchies relating simple concepts and skills to more complex ones.

In this particular module, Stat Lady's instruction consists of five sections: (a) frequency distributions, (b) proportions and percentages, (c) grouped frequency distributions, (d) cumulative frequency distributions, and (e) plotting. There is a decidedly experiential flavor to the instruction, during which learners are required to participate in the interactive lessons. To illustrate, Stat Lady begins the section on proportions by displaying a frequency distribution of the ages of 20 children who visited a pediatric clinic on a particular day. Stat Lady summarizes information about the distribution and makes the case that it is often useful to represent data in the form of proportions, "where a *proportion* is a ratio involving a specific value divided by the total." This last statement becomes more concrete when it is joined on the screen by an animated pie that has a slice moving into and out of it. Stat Lady continues the instruction by relating the pie slice to the part of the age distribution that represents 10-year-old children: "Two out of 20 children are 10 years old. The phrase '2 out of 20' can be converted to a proportion. What is the general formula for computing proportions?"

At this point, the learner begins participating in the instructional lesson by constructing an answer to the question from an on-line equation-builder tool that contains about 10 different symbols (e.g., Σ, N, X, f, *, +, /). The program "knows" that the correct answer is f/N (or f/Σf) and thus can address specific errors accordingly. For learners who answer the question correctly the first time (with no assistance from Stat Lady), the program congratulates them with a comment such as, "Isn't that special? You figured out that the formula for computing proportions is f/N." Feedback also addresses specific incorrect responses. For instance, if a learner enters Σf as the answer, Stat Lady would say, "This notation stands for summing the frequency column. Please try again." In all cases, if a student provides an incorrect answer, Stat Lady intervenes with progressively more specific feedback related to the particular error. On any given problem, learners are allowed up to three errors before Stat Lady provides the correct answer. Even then, the student has to enter the correct answer before being allowed to continue.

Each instructional section is followed by a set of problem-solving scenarios that test the student's abilities on the material presented in the preceding instruction. To illustrate using a proportions example, learners begin by choosing a scenario from a pool of five topics (e.g., company supervisor problem concerning absenteeism rates, sociology instructor's research on dating habits of college students). Within each scenario, the student demonstrates skills and answers questions related to each CE, and the tutor assesses performance. In this example, a learner creates a frequency distribution and computes proportions using tutor-supplied data relating to college student dating habits. After the table is created, the student answers specific questions about the distribution, such as, "What is the most frequently occurring value in your distribution?" "Do any values in your distribution not occur?" "Σp is always equivalent to what value?" and so forth. Each scenario is similar to others in the problem set because the questions assess the same group of CEs; however, the data and context of each scenario differ.

At any time and regardless of condition, learners are free to open the on-line dictionary, the formula bank, or both, to see definitions of terms and proper formulas. The dictionary contains full descriptions of the concept or rule in question, constructed in hypertext (i.e., one can click on italicized terms). For example, the entry for proportion states, "A proportion (p) describes the *frequency* of some occurrence (f) in relation to the total number (N). Proportions can be expressed as fractions or decimals, computed as: $p = f/N$. The sum of all proportions in a distribution equals 1 ($\Sigma p = 1$)." The formula bank consists primarily of formulas and graphs.

Other materials. In addition to the tutor itself, other materials used in this study included an on-line demographic questionnaire, a pretest, a posttest, and an affective survey. Items on the demographic questionnaire assessed gender, age, computer and video game experience, general educational background (years of school completed), and previous mathematics education (e.g., how many prior algebra, geometry, and calculus courses completed). The on-line pretest contained duplicate items for every CE covered in the tutor to provide a more accurate representation of incoming domain-related ability level. Like the practice items within the tutor, these items assessed symbolic knowledge (e.g., selecting definitions of concepts and providing notation, formulas, or both), procedural skills (e.g., computing a percentage, creating frequency distribution tables and graphs), and conceptual knowledge (e.g., interpreting frequency distribution tables and graphs). Posttest items were isomorphic to those in the pretest, enabling us to examine learning gains. The affective survey asked learners how they felt about the tests and the tutor (e.g., "How much did you feel like you were in control over the pace at which you were learning?" "How frustrating was the tutor?"), presented in a 1–7 Likert-scale format. The survey had a total of 10 items and was administered following the posttest.

Design and Procedure

Participants were randomly assigned to one of the five practice conditions and were tested in groups of about 20. There were four fixed-practice conditions, with one practice opportunity defined as one scenario. In our study, we defined abbreviated practice as solving one scenario per section and extended practice as solving four scenarios per section. Recall that there were five sections altogether in this module.

Two of the practice conditions involved switching participants from one condition to the other after the third section. This resulted in the mixed-practice conditions: abbreviated-extended (AE) and extended-abbreviated (EA). The other two conditions held the number of practice opportunities constant: abbreviated-abbreviated (AA) and extended-extended (EE). The fifth condition was the LC condition, in which each learner completed one to four scenarios per section at their discretion. Thus, there was considerable variability in the number of scenarios completed in each condition. For instance, individuals in the AA condition would have to solve only one scenario for each of the five problem sets (for a total of five scenarios), whereas participants in the EE condition would solve four scenarios per section, for a total of 20 scenarios. After completing both the questionnaire and the pretest, participants proceeded to learn from Stat Lady. Upon completion of the tutor, the on-line posttest and affective survey were administered.

Participants were asked to return six months later to take part in the follow-up portion (Phase 2) of the study. About one third (n = 120) of the original participants returned for retention testing. The average lag between short-term and long-term retention testing was 29.2 weeks (SD = 6.2 weeks). The number and relative proportion of returning participants by condition was AA (n = 24, 28%), AE (n = 19, 32%), EA (n = 21, 36%), EE (n = 24, 27%), and LC (n = 32, 36%). These values differ only by chance.

Testing for the retention part of the study was conducted in small groups of about five persons. Test administrators briefed each group on the importance of trying to remember as much as possible from the original learning session. The test items themselves were isomorphic to those from the first half of the study in that they were of the same structure but involved new problems.

RESULTS

Prior to making comparisons among practice conditions in terms of our dependent measures, we needed to ensure that learners within each condition were demographically comparable. The sample sizes per condition were AA (n = 86), AE (n = 60), EA (n = 58), EE (n = 88), and LC (n = 88). Several one-way analyses of variance (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.

TABLE 1: Summary of Pretest, Posttest, and Six-Month Test Scores by Condition

	AA	AE	EA	EE	LC
Pretest					
Mean	40.29	43.32	46.25	46.34	43.76
SD	14.12	14.94	14.90	13.92	15.40
Min/Max	19/82	19/86	16/87	20/76	20/83
Posttest					
Mean	65.34	69.74	75.06	78.17	71.86
SD	20.20	18.29	17.60	16.25	18.69
Min/Max	25/99	30/98	35/99	35/99	24/99
Six-month test					
Mean	49.92	62.41	59.57	67.66	61.85
SD	18.82	19.99	15.43	16.21	20.43
Min/Max	21/89	33/95	33/87	35/93	32/94

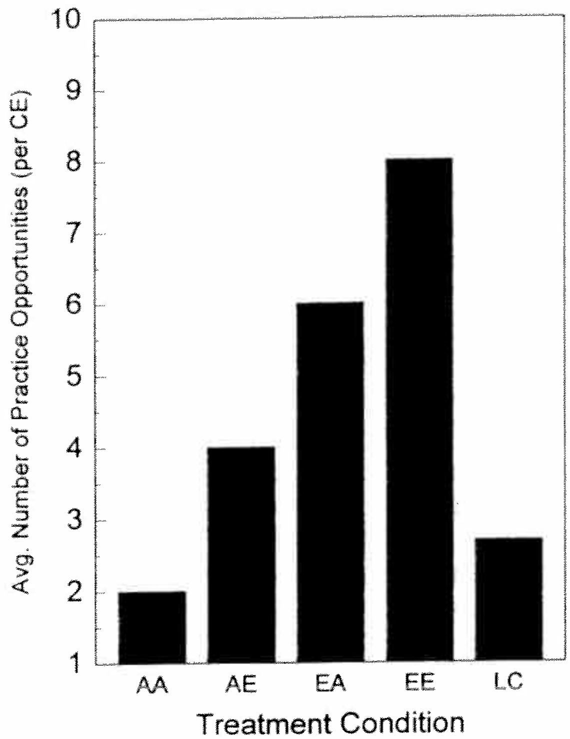

Figure 1. Average number of problems solved (per CE) by treatment condition.

We also examined pretest scores by condition to see if there were incoming differences in statistics knowledge. (We expected low variability on this measure because we solicited participants with minimal statistics backgrounds and made random assignments of learners to condition.) There were marginally significant differences on pretest scores across conditions, $F(4, 375) = 2.32$, $p = .06$. The means and range of pretest scores by condition are shown in the top row of Table 1.

Amount of Practice

We computed an ANOVA on the number of problems requested and solved as the dependent variable and practice condition as the independent variable. It was hardly surprising that we found significant differences among our practice conditions in terms of the amount of practice variable, given that four of the conditions had a fixed number of practice opportunities, $F(4, 375) = 791.01$, $p < .001$. What was uncertain was the amount of practice that LC learners would impose on themselves relative to the fixed-practice conditions. We hypothesized an intermediate number of problems solved by this group, representing an average of those choosing few, many, and in-between. Unexpectedly, we found that the LC participants almost uniformly choose to solve very few problems, midway between participants in the AA and AE conditions. Figure 1 shows the average number of practice opportunities per curriculum element by condition.

We computed individual *t* tests comparing participants in the LC condition with participants in the AA and AE conditions and found these samples to be significantly different: LC versus AA $t(172) = -5.07$, $p < .001$; LC versus AE $t(146) = 6.107$, $p < .001$. Figure 2 displays the distribution of number of problems that the LC participants elected to complete – this was a highly skewed distribution, illustrating that the majority of LC individuals elected to solve a minimal number of problem scenarios.

Learning: Short-Term and Long-Term Gain

To test for immediate and long-term differences in knowledge and skill acquisition, we created two different gain scores. The first dependent variable (short-term gain) assessed the degree to which the participants acquired the new material from pretesting to immediate testing (i.e., posttest score minus pretest score). The second dependent variable (long-term gain) assessed the overall degree of improvement from pretesting to testing after six months had elapsed (i.e., six-month test score minus pretest score). Figure 3 shows the pretest, posttest, and six-month test data for each of the five conditions in the study. Table 1 summarizes the means, standard deviations, and range of scores on these variables by condition.

Short-term gain. Do individuals differ in their acquisition of new knowledge and skills, and if so, is it affected by their training condition (amount of practice), incoming knowledge level, or both? We computed an ANOVA on our short-term gain variable by condition (i.e., AA, AE, EA, EE, LC) and level of domain-specific incoming knowledge (i.e., low vs. high), computed as a median split of pretest data across the entire sample of participants. We included level of incoming knowledge as an independent variable in order to examine the data for possible aptitude-treatment interactions. That is, some learners (e.g., low

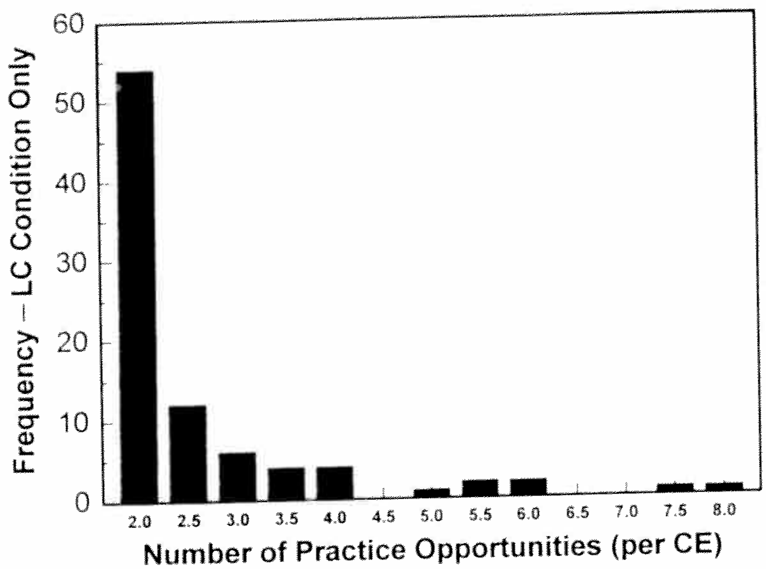

Figure 2. Number of practice opportunities (per CE) for the LC group only.

incoming knowledge) might show optimal gain having learned under extended practice conditions, whereas others (e.g., high incoming knowledge) might show greater gain from the LC condition.

Unlike the results reported by Shute and Gawlick (1995), the ANOVA showed a significant main effect on short-term gain attributable to condition, $F(4, 370) = 4.11$, $p < .01$. The ordering of learning gain increased relative to the amount of practice: AA (M = 25.04) < AE (M = 26.42) < LC (M = 28.10) < EA (M = 28.82) < EE (M = 31.83). Learners in the LC condition showed intermediate gain relative to the other practice conditions. Post hoc comparisons were computed between LC learners and those in each of the fixed-practice conditions, and there were no significant mean differences (Bonferroni tests). There were, however, significant differences in gain scores between learners in the most extended (EE) and those in the abbreviated (AA and AE) conditions (effect sizes = 0.61, $p < .001$, and effect sizes = .46, $p < .03$, respectively).

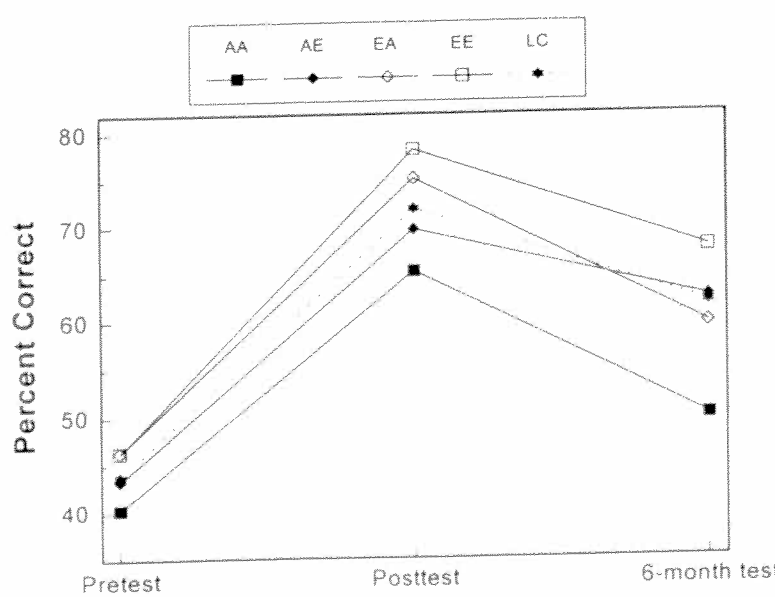

Figure 3. Pretest, posttest, and six-month test scores by condition.

The main effect of incoming knowledge level on short-term gain was significant, $F(1, 370) = 7.62$, $p < .001$, for which learners with low incoming knowledge ($M = 26.29$, $SD = 12.44$) demonstrated lower gain scores relative to learners with high incoming knowledge ($M = 29.94$, $SD = 8.46$). The interaction between condition and incoming knowledge was not significant, $F(4, 370) = 2.03$, $p = .09$.

Long-term gain. After six months, do individuals differ in their retention of knowledge and skills acquired earlier? If so, is retention affected by original practice condition, initial knowledge level, or both? Before making comparisons among conditions on retention, we first needed to ensure that the subset of returning learners was comparable to the original sample (Phase 1 data overall and per practice condition). We computed one-way ANOVAs on demographic measures (age, gender, education, and computer experience) by phase. None of these measures were significantly different. We also compared returning participants' data with original participants' data on Phase 1 posttest scores overall and by condition. Scores from the returning sample did not differ significantly from the original sample on this measure, overall $t(454) = -0.11$, $p = .91$.

We then computed an ANOVA on our long-term gain variable (i.e., retention) by condition and level of domain-specific incoming knowledge (low vs. high). Unlike findings from the short-term gain analysis, there was no main effect of practice condition on long-term gain, $F(4, 110) = 1.14$, $p = .34$. In addition, the main effect attributable to incoming knowledge level was not significant, $F(1, 110) = 3.58$, $p = .06$, and neither was the interaction between condition and incoming knowledge, $F(4, 110) = 2.37$, $p = .06$.

Learning Time and Efficiency Measures

Tutor-learning time. We decomposed the total tutor time variable into two parts – instruction time and problem-solving time – reflecting the two distinct stages in the Stat Lady program. Instruction time should vary in relation to one's facility in acquiring and understanding the new material, whereas problem-solving time should vary in relation to practice condition. Three ANOVAs were computed on instruction time, problem-solving time, and total time required to complete the tutor. All three variables showed significant differences attributable to condition. Shorter amounts of instructional time were needed for participants who had the benefit of being in the more extended practice conditions, EE < EA < LC < AE < AA. Problem-solving time showed the opposite pattern, increasing as a function of amount of practice from least to most, AA < LC < AE < EA < EE. The ordering of conditions by the total-time variable was unexpected, LC < AA < AE = EA < EE. Table 2 shows the time data for each of these measures with the respective *F* values.

We computed post hoc comparisons among conditions on the total time variable and found that learners in the AA condition required significantly less time than learners in the other fixed conditions, $t(375) = -3.56$, $p < .001$, effect size = −0.42. Similarly, learners in the LC condition required significantly less time to complete the tutor than did those in the other conditions, $t(375) = -4.33$, $p < .001$, effect size = −0.53, but did not specifically differ from the AA learners, $t(375) = -0.48$, $p = .63$. The largest contrast was between the LC and EE learners, $t(375) = -6.17$, $p < .001$,

TABLE 2: Instruction, Problem Solving, and Overall Tutor Time (h) by Condition

Condition	Instruction	Problem Solving	Total
AA (n = 86)	2.05	1.98	4.03
AE (n = 60)	1.88	2.96	4.84
EA (n = 58)	1.69	3.16	4.85
EE (n = 88)	1.52	3.95	5.47
LC (n = 88)	1.78	2.13	3.91
F (df = 4, 375)	8.78	40.38	12.74
p	< .001	< .001	< .001

Note: A = abbreviated, E = extended, and LC = learner control practice conditions.

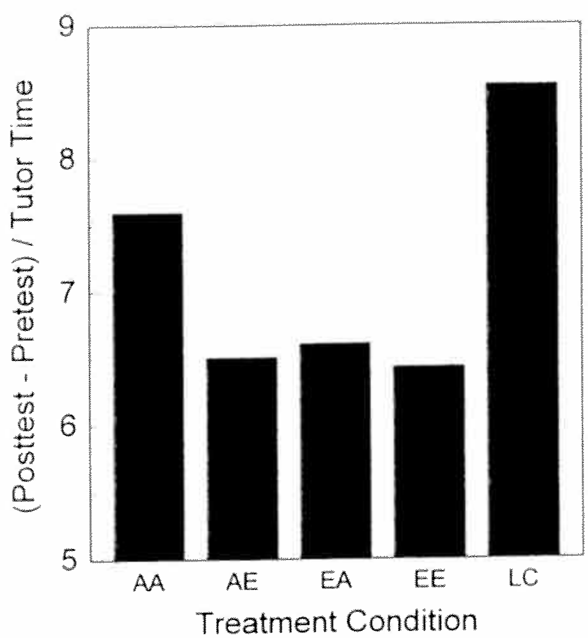

Figure 4. Effect of condition on short-term efficiency index.

effect size = –.90, with LC learners almost one standard unit faster than the EE learners overall.

Short-term efficiency. To make comparisons among conditions in terms of relative efficiency, we first computed a ratio of learning gain (i.e., difference between pretest and immediate test scores) to learning time (i.e., total time learning from tutor) for each learner. Larger values for this index are interpreted as greater efficiency. The distribution of the overall sample on this variable ranged from 0.18 to 26.33 ($M = 7.22$, $SD = 6.5$).

We computed an ANOVA on this ratio to test for differences in short-term learning efficiency as a function of condition and incoming knowledge level (again, designated as low vs. high depending on whether the learner was below or above a median split on pretest data). There was a significant main effect of condition on this variable, $F(4, 370) = 6.12$, $p < .001$, shown in Figure 4. Learners assigned to the LC condition showed significantly greater efficiency indices compared with learners in the AE (effect size = 0.50), EA (effect size = 0.65), and EE (effect size = 0.77) conditions. The associated post hoc comparisons (Bonferroni tests) were significant: LC versus AE, $p = .03$; LC versus EA, $p = .05$; LC versus EE, $p = .01$. Learners in the LC and AA conditions were not significantly different from each other, and AA learners did not differ significantly from the other fixed-practice conditions. The main effect of incoming knowledge was significant, $F(1, 370) = 111.64$, $p < .001$. Learners with low incoming knowledge showed poorer short-term efficiencies compared with learners with high incoming knowledge.

Finally, the interaction between condition and incoming knowledge was significant, $F(4, 370) = 5.35$, $p < .001$. These data are shown in Figure 5, and the graph suggests that

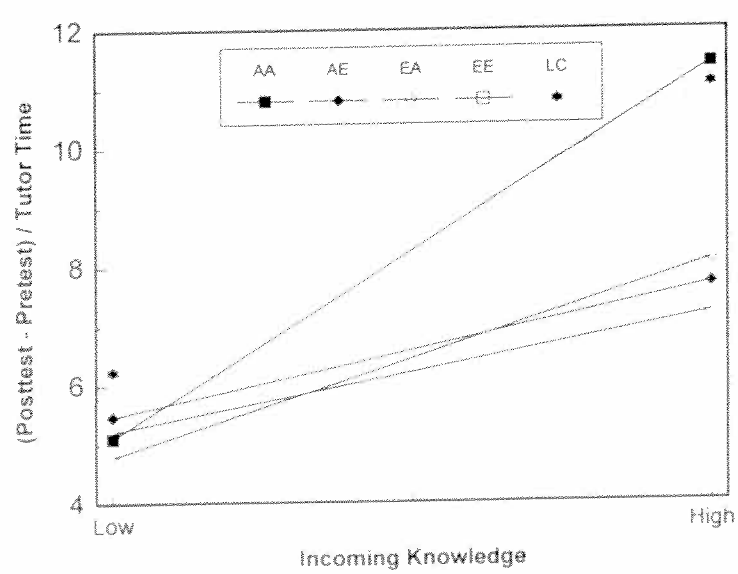

Figure 5. Interaction between condition and prior knowledge on short-term efficiency index.

for both high and low incoming knowledge learners, placement in the LC condition enhanced short-term efficiency. However, when one looks only at the learners categorized as having low incoming knowledge, the participants in the LC condition were only marginally more efficient than those in the fixed-practice conditions (AA, AE, EA, EE), $t(185) = 1.81$, $p = .07$, effect size = 0.33. A more striking contrast is seen with the high incoming knowledge learners. In this case, the high incoming knowledge learners assigned to the LC condition were significantly more efficient than their counterparts in the other fixed-practice conditions (AE, EA, EE), $t(185) = 5.41$, $p < .001$, effect size = 1.21. The contrast between high incoming knowledge learners in the LC and AA conditions was not significant.

Long-term efficiency. We computed a long-term efficiency index similar to the short-term efficiency index described previously. The numerator consisted of long-term learning gain (i.e., the difference between pretest score and six-month test score) divided by time spent originally learning from the tutor. It is interpreted in the same manner as the short-term gain index, with larger numbers associated with greater efficiency. The distribution of the overall sample on this variable ranged from 0.16 to 22.89 ($M = 4.51$, $SD = 3.6$).

We computed an ANOVA on the long-term efficiency variable to test for differences in long-term learning efficiency as a function of original practice condition and incoming knowledge level. Our independent variables were condition and pretest score – low versus high depending on whether the learner was below or above a median split on pretest data.

There was a significant main effect of condition on this variable, $F(4, 110) = 4.93$, $p = .001$. These data are shown in Figure 6. Overall, learners in the LC condition demonstrated much greater efficiency scores than those in the other fixed-practice conditions (AA, AE, EA, EE): $t(115) = 3.34$, $p = .001$, effect size = 0.90. There was also a significant main effect of incoming knowledge on long-term efficiency, $F(1, 110) = 43.48$, $p < .001$, with an advantage shown by the high incoming knowledge learners. Finally, the interaction between incoming knowledge and condition

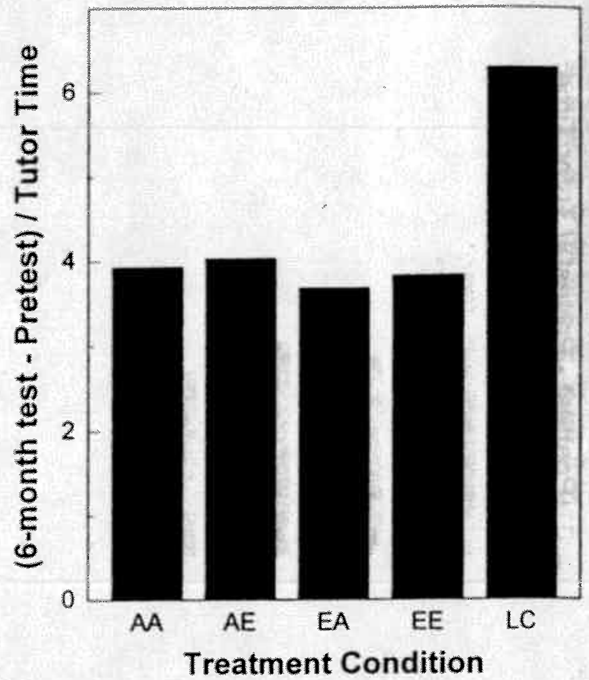

Figure 6. Effect of condition on long-term efficiency index.

on long-term efficiency was significant, $F(4, 110) = 4.86$, $p = .001$. These data are shown in Figure 7 and suggest that learners with low incoming knowledge did not appear to benefit from any particular learning condition (i.e., there were no discernable differences among conditions). However, learners with relatively high incoming knowledge showed the greatest long-term efficiency if they originally learned from the LC condition.

Planned comparisons were computed separately for the low and high incoming knowledge participants. Findings revealed no significant differences among the low incoming knowledge participants in terms of practice condition (individual Bonferroni tests were not significant for any of these comparisons). With regard to the high incoming knowledge participants, those in the LC condition were significantly more efficient than their counterparts in the fixed-practice conditions (AE, EA, EE), $t(50) = 4.51$, $p < .001$, effect size = 2.2. The individual Bonferroni tests were all significantly different, LC versus AE effect size = 1.57, $p = .02$; LC versus EA effect size = 2.44, $p = .05$; LC versus EE effect size = 2.87, $p < .001$. The contrast between high incoming knowledge learners in the LC and AA conditions was not significant.

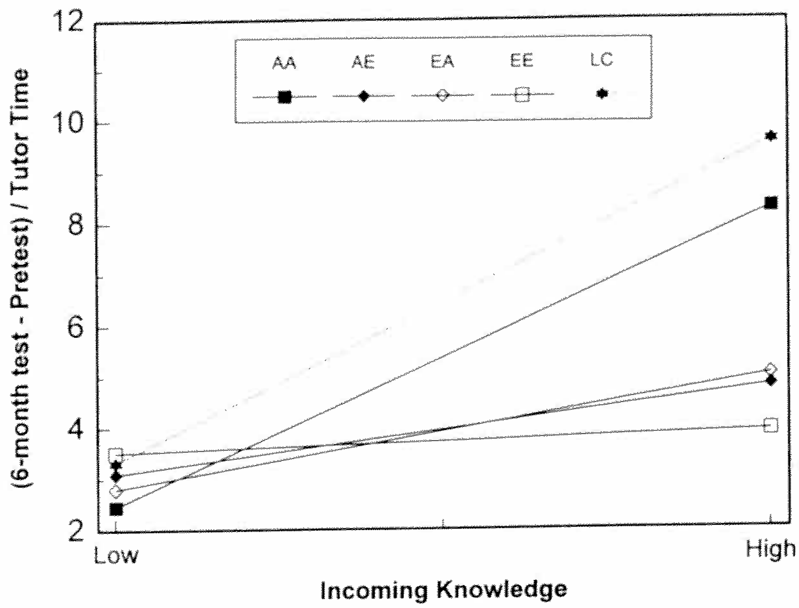

Figure 7. Interaction between condition and prior knowledge on long-term efficiency index.

DISCUSSION

The experiment described in this paper explores the effects of differential practice opportunities and learner control on short- and long-term learning gains and efficiency. We examined two global research issues: the replication of anomalous findings (reported by Shute & Gawlick, 1995) and the investigation of learner control on learning outcome and efficiency.

Regarding the first issue (and disregarding the learner control manipulation for the moment), we did not replicate the null differences in short-term gain scores across practice conditions that were reported in the Shute and Gawlick (1995) study. Instead, we found a pattern of differences among practice conditions that lends additional support to the "practice makes perfect" (or at least "practice makes better") position, because learners given the most practice opportunities showed the greatest learning gains compared with learners with fewer practice opportunities. However, as is typically found in the literature on practice effects, the increased outcome levels resulting from extended practice occur at the expense of efficiency (i.e., more practice results in greater learning time).

We also did not replicate the anomalous finding of greater long-term retention of knowledge and skills by those who originally learned from mixed-practice conditions (AE, $n = 10$, and EA, $n = 6$). Shute and Gawlick (1995) predicted that those who originally learned from the more extended condition (EE, $n = 5$) would show greater long-term retention. In the present study, following six months between original instruction and retention testing, we found no main effect of condition on degree of retention (i.e., long-term gain). One possible reason for this replication failure is that the sample sizes for the mixed-group conditions in the 1995 study were very small (attributable to the two-year lag between posttest and retention test). This suggests a power issue, and our current sample size is larger, thus more robust. Also, the criterion task in the earlier study was much longer than the current one (i.e., 12.0 h vs. 5.5 h on average for the EE conditions, respectively). Consequently, fatigue could easily have been a factor in the earlier study, decreasing the relative efficacy of the condition for those who could have benefited most from it (i.e., individuals with low incoming knowledge). We conclude that the effects of practice are, indeed, apparent immediately following the learning task and tend to attenuate over time.

The more exciting results from this study relate to the learner-control manipulation.

When this condition was compared with fixed-practice conditions on several dependent measures, there were a few surprises. First, we found that learners in the LC condition chose minimal amounts of practice, on par with the two most abbreviated practice conditions. Furthermore, the LC participants required the least amount of time to complete the tutor compared with participants in the fixed-practice conditions. However, despite their abbreviated practice schedule, they still managed to exhibit immediate gain scores that are comparable to those shown by learners in the most extended practice condition. As a result, when we examined efficiency indices (relating gain score to tutor-learning time), the LC group showed significantly greater efficiencies for both short- and long-term gains compared with the fixed-practice conditions. With respect to only the long-term efficiency variable (see Figure 6), LC individuals outperformed the fixed-practice conditions by approximately one standard deviation. These findings indicate that learners should be given some degree of control over their practice opportunities, particularly when the instructional emphasis is on learning efficiency.

When incoming knowledge of our learners is considered, we found interactions with practice condition for both short- and long-term efficiency measures. Learners with low incoming knowledge did not differ in terms of short- or long-term efficiency by condition. That is, their efficiency indices were comparable across the five practice conditions for both immediate and delayed testing. In contrast, participants with high incoming knowledge who were in the LC condition scored more than two standard deviations above three of the four fixed-practice conditions on our long-term efficiency index (see Figure 7). These interactions are similar to findings in the literature in which high-ability learners and those possessing greater levels of prior subject matter understanding benefited most from LC treatments (e.g., Gay, 1986; Kinzie, 1990). The literature also suggests that persons with low ability (or less background knowledge) perform better when the computer has control (e.g., Lee & Lee, 1991; Lee & Wong, 1989). However, we failed to find evidence of this in the current study.

IMPLICATIONS FOR TUTORING ENVIRONMENTS

Overall, these results appear promising for supporters of learner control, and one might be tempted to conclude that it is generally better to allow for learner control over practice opportunities than not to make that provision. We want to be cautious, however, in advocating any broad generalizations. Let us take a moment to consider in more detail the behavior of our LC participants and also the particulars of this tutoring environment as we develop the reasons for our caution.

Looking at the short- and long-term learning efficiency results displayed in Figures 5 and 7, there were no significant differences in learning efficiency between the LC and the AA participants. Furthermore, learners in both of these groups were more efficient than those in the other fixed-practice conditions. Note also that we found no significant differences between LC and AA participants in terms of learning gain. However, we report a main effect of practice such that more practice produced better short-term gain across all conditions. Thus, we have a tutoring situation in which more practice is better than less practice with respect to short-term gain, but less practice is better than more in terms of efficiency.

This pattern of results is attributable to a combination of the ubiquitous power law of learning and the particular construction of this tutor. The power law states that although one can expect some added benefit, there will be diminishing educational returns with each additional practice opportunity. At some point in the learning process, it no longer pays to keep practicing; the additional practice does not provide sufficient return. If those practice opportunities are costly (i.e., take a long time), if the material can be learned quickly, or both, then it would be inefficient to complete more than a minimum number of problems.

With regard to this particular tutor, two independent studies provide evidence for both of those conditions in Stat Lady. Shute (1995) elaborated on the fact that each additional scenario is fairly time consuming, given that every curriculum element in a section is tested at least once in each scenario. That establishes the cost factor. Gluck (1997) provided evidence for the possibility of relatively early learning. Based on performance scores and verbal protocol data from four high incoming knowledge participants (learning from the same Stat Lady module as was employed in this study), he concluded that most of their learning events took place by the end of the first scenario.

We conclude that this is roughly the situation in which the high incoming knowledge participants in this study found themselves. They were aware that they were learning the material fairly quickly and that there was a high cost to doing more scenarios, so they generally opted out after the first one. Of course, we are giving them the benefit of the doubt with respect to their self-monitoring abilities. An alternative explanation is that they were simply lazy and did not want to complete any more problems after the first scenario and that the fact that they were learning the material quickly was incidental. We prefer the self-monitoring explanation to the laziness explanation but are unable to rule out either one conclusively on the basis of the available data.

We have shown that allowing for learner control over practice opportunities has the potential to result in efficient learning. What this means for instructional design, however, is not yet clear. What we have done so far is to couch the interpretation of our results in a multidimensional framework. This framework includes factors such as characteristics of the learners (e.g., incoming knowledge), characteristics of the learning environment (e.g., cost of completing a problem scenario), and the instantiation of the domain in that learning environment (e.g., learnable relatively quickly). In order to arrive at recommendations for instructional design, one needs to add some explicit statement regarding the goal of the instruction. For instance, if the goal of the instruction was to ensure that the average posttest score would be above 75% (and time was not an issue), then computer control would be the recommendation. If the goal was to provide the most efficient long-term learning gain ("the most bang for the buck") and only the instructional options available in this study were given, then learner control is the clear winner.

ACKNOWLEDGMENTS

We would like to thank Barry Goettl, Patrick Kyllonen, Lisa Torreano, Rickard Robbins, Cathy Connolly, and Shirley Snooks for their varied and helpful contributions to this study. The research reported in this paper was conducted as part of the U.S. Air Force Research Laboratory TRAIN Project, Brooks Air Force Base, Texas. 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 U.S. Air Force.

REFERENCES

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

Belland, J. C., Taylor, W. D., Canelos, J., Dwyer, F., & Baker, P. (1985). Is the self-paced instructional program, via microcomputer-based instruction, the most effective method of addressing individual differences? *Educational Communications and Technology Journal, 33*(3), 185–198.

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

Carrier, C. A. (1984). Do learners make good choices? *Instructional Innovator, 29*(2), 15–17.

Dalton, D. W. (1990). The effects of cooperative learning strategies on achievement and attitudes during interactive video. *Journal of Computer-Based Instruction, 17*(1), 8–16.

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted video instruction. *Journal of Educational Psychology, 78*, 225–227.

Gluck, K. A. (1997). *The adaptive nature of learning from Stat Lady*. Unpublished master's thesis, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Goetzfried, L., & Hannafin, M. J. (1985). The effect of locus of CAI control strategies on the learning of mathematics rules. *American Educational Research Journal, 22*(2), 273–278.

Gray, S. H. (1987). The effect of sequence control on computer assisted learning. *Journal of Computer-Based Instruction, 14*(2), 54–56.

Hannafin, M. J. (1984). Guidelines for using locus of instructional control in the design of computer-assisted instruction. *Journal of Instructional Development, 7*(3), 6–10.

Higginbotham-Wheat, N. (1988, November). *Perspectives on implementation of learner control in CBI*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Lexington, KY. (ERIC Document Reproduction Service No. ED 305 898)

Hurlock, R. E., Lahey, G. F., & McCann, P. H. (1974). *Student-controlled versus program controlled CAI*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Johansen, K. J., & Tennyson, R. D. (1983). Effect of adaptive advisement on perception in learner controlled, computer-based instruction using a rule-learning task. *Educational Communications and Technology Journal, 31*(4), 226–236.

Kinzie, M. B. (1990). Requirements and benefits of effective interactive instruction: Learner control, self-regulation, and continuing motivation. *Educational Technology Research and Development, 38*(1), 5–21.

Kinzie, M. B., & Sullivan, H. J. (1989). Continuing motivation, learner control, and CAI. *Educational Technology Research and Development, 37*(2), 5–14.

Lahey, G. F., Hurlock, R. E., & McCann, P. H. (1973). *Post lesson remediation and student control of branching in computer-based training* (Report SRR-73-19). San Diego, CA: Naval Personnel Research and Development Center. (ERIC Document Reproduction Service No. ED 083 797)

Lee, S. S., & Lee, Y. H. K. (1991). Effects of learner-control versus program-control strategies on computer-aided learning of chemistry problems: For acquisition or review? *Journal of Educational Psychology, 83*, 491–498.

Lee, S. S., & Wong, S. C.-H. (1989). Adaptive program vs. learner control strategy on computer aided learning of gravimetric stoichiometry problems. *Journal of Research on Computing in Education, 21*, 367–379.

Lepper, M. R. (1985). Microcomputers in education: Motivational and social issues. *American Psychologist, 40*(1), 1–18.

MacGregor, S. K. (1988). Instructional design for computer-mediated text systems: Effects of motivation, learner control, and collaboration on reading performance. *Journal of Experimental Education, 56*(3), 142–147.

Merrill, M. D. (1975). Learner control: Beyond aptitude-treatment interactions. *AV Communications Review, 23*, 217–226.

Merrill, M. D. (1980). Learner control in computer based learning. *Computers & Education, 4*, 77–95.

Murphy, M. A., & Davidson, G. V. (1991). Computer-based adaptive instruction: Effects of learner control on concept learning. *Journal of Computer-Based Instruction, 18*(2), 51–56.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–56). Mahwah, NJ: Erlbaum.

Relan, A. (1991). Effects of two levels of strategy training and learner control on performance and choice behavior in a computer-based concept-classification task (Doctoral dissertation, University of Minnesota, 1991). *Dissertation Abstracts International, 52*, 3895-A.

Rivers, L. C. (1972). *Development and assessment of an adaptive strategy utilizing regression analysis for the presentation of instruction via computer* (Tech. Report 27). Tallahassee, FL: Florida State University, Computer-Assisted Instruction Center. (ERIC Document Reproduction Service No. ED 077 228)

Ross, S. M., & Morrison, G. R. (1989). In search of a happy medium in instructional technology research: Issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development, 37*(1), 19–33.

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: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

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*, 781–803.

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

Shyu, H.-Y., & Brown, S. W. (1992). Learner control versus program control in interactive videodisc instruction: What are the effects in procedural learning? *International Journal of Instructional Media, 19*(2), 85–96.

Steinberg, E. R. (1989). Cognition and learner control: A literature review, 1977–1988. *Journal of Computer-Based Instruction, 16*(4), 117–121.

Tennyson, R. D. (1980). Instructional control strategies and content structure as design variables in concept acquisition using computer-based instruction. *Journal of Educational Psychology, 72*, 525–532.

Tennyson, R. D., & Buttrey, T. (1980). Advisement and management strategies as design variables in computer-assisted instruction. *Educational Communications and Technology Journal, 28*(3), 169–176.

Tennyson, R. D., Park, O.-C., & Christensen, D. L. (1985). Adaptive control of learning time and content sequence in concept learning using computer-based instruction. *Journal of Educational Psychology, 77*, 481–491.

Tennyson, C. L., Tennyson, R. D., & Rothen, W. (1980). Content structure and instructional control strategies as design variables in concept acquisition. *Journal of Educational Psychology, 72*, 499–505.

Tennyson, R. D., Welsh, J. C., Christensen, D. L., & Hajovy, H. (1985). Interactive effect of information structure, sequence of information, and process learning time on rule learning using computer-based instruction. *Educational Communications and Technology Journal, 33*(3), 213–233.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs, 2*(Whole No. 8).

Williams, M. D. (1996). Learner-control and instructional technologies. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 957–983). New York: MacMillan.

Valerie J. Shute received her Ph.D. in 1984 from the University of California, Santa Barbara, in educational/cognitive psychology. Since 1986 she has been employed at the U.S. Air Force Research Laboratory, Brooks Air Force Base, Texas.

Lisa A. Gawlick received her B.A. in psychology in 1989 from St. Mary's University in San Antonio, Texas. She is a senior research assistant at Galaxy Scientific Corporation in conjunction with the U.S. Air Force Research Laboratory at Brooks Air Force Base, Texas.

Kevin A. Gluck earned his M.S. in cognitive psychology in 1997 at Carnegie-Mellon University, where he is pursuing his doctorate in cognitive psychology.

Date received: February 25, 1997
Date accepted: December 3, 1997