Deconstructing a Computer-Based Tutor: Striving for Better Learning Efficiency in Stat Lady

Kevin A. Gluck[1], Valerie J. Shute[2], John R. Anderson[1], and Marsha C. Lovett[1]

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15217, USA
{kgluck, ja, lovett}@cmu.edu
[2]Air Force Research Laboratory/HEJT, 1880 Carswell Avenue, Lackland AFB, TX 78236, USA
vshute@colab.brooks.af.mil

Abstract. This paper is a prospective report of an ongoing design project involving a computer-based tutoring system called Stat Lady (Shute & Gluck, 1994). Previous studies have shown considerable improvement on curriculum objectives as a result of interation with the tutor. The goal now is to try to improve on learning efficiency, defined as knowledge gain per unit of time. The question we will be asking in this study is: What can we peel away from the tutor to make it a more efficient teaching tool, without having a negative impact on curriculum learning? Additionally, as we remove pieces of the tutor, what effect (if any) will that have on the subjective enjoyment of the learning experience? The study in progress investigates these issues in a 2x2 factorial design varying the presence or absence of contextualized instruction and problem solving across the presence or absence of certain interface features that were an integral part of the original tutor.

Introduction

> „That with psychological attention to human performance airplanes became more flyable encourages us to believe that with psychological attention to human performance computers can become more usable."
>
> Card, Moran, and Newell (1983, p. 9)

Card, Moran, and Newell (1983) penned those hopeful words more than 15 years ago in the introductory chapter to their book on the application of information-processing psychology to the study of human-computer interaction. In the twelfth chapter of their text, Card et al. outline a framework for applying psychology in this important domain, characterizing it as an open design process, such that the computer program is regularly redefined all along the way. They write, „Thus, design proceeds in a complex, iterative fashion in which various parts of the design are incrementally generated, evaluated, and integrated" (p. 406).

Koedinger and Anderson (in press) express a similar opinion in their forthcoming paper on principled design of cognitive tutors. They note that, „It should be the natural expectation of the field that no Interactive Learning Environment will be fully effective in its intial implementations and that early demonstrations of limitations have a positive, not a negative, bearing on the value of the final system." This is to be expected, of course, because of the combinatorial complexity at the intersection of the characteristics of human users, potential variations in the characteristics of computer-based instructional systems, and differences in the pedagogical requirements of different domains.

This paper is a prospective[1] report of a design project involving a computer-based tutoring system called Stat Lady (Shute & Gluck, 1994), which teaches introductory descriptive statistics. In contrast to ITS intended for classroom deployment, Stat Lady was originally developed strictly as a tool for student modeling research in the Armstrong Laboratory TRAIN facility. Satisfyingly, however, and despite the fact that its *raison d'etre* is to be a piece of basic research software, it has been demonstrated that Stat Lady results in a good deal of curriculum-related learning on the part of those who work through it, and laboratory subjects report actually enjoying the experience (Shute, 1995). The question currently under investigation is this: What can we peel away from the tutor to make it a more efficient teaching tool, without having a negative impact on curriculum learning? Additionally, as we remove pieces of the tutor, what effect (if any) will that have on the subjective enjoyment of the learning experience? Participants in previous studies have reported that they enjoyed learning from the tutor. If altering it destroys the positive affective reaction to the experience, thereby decreasing student interest in pursuing future opportunities to learn about the domain (or in learning about a different domain from a new Stat Lady instructional module), then any gain in learning efficiency could ultimately be overshadowed by losses in motivation.

1.1 The Research Goal

Implicit in the above description of our current research question is that the goal is to increase *learning efficiency*, specifically by improving on the speed with which learners can use the tutor. Learning efficiency, defined as the amount of knowledge acquired per unit of time, is of paramount concern in the educational and training communities. There are two ways to improve on learning efficiency. One can either find a way to get students to (a) learn more - without increasing their time investment, or get them to (b) learn faster - without decreasing their learning.

[1]As of the due date for ITS '98 conference submissions, we are in the process of redesigning the tutor. Subjects will be run in the coming months, and the data will be ready for presentation by August.

The goal of increasing learning efficiency is the driving motivation behind a great deal of research in education and related fields. The ITS community is one of those fields. In this particular project, we are attempting (b) rather than (a). The primary reason for this is that Gluck, Lovett, Anderson, and Shute (1998) recently completed a study investigating how people use Stat Lady and how interaction with the tutor changes with practice. Through the analysis of concurrent verbal protocols (see Ericsson & Simon, 1993) collected as participants learned from Stat Lady, as well as conclusions derived from post-participation interviews, it became apparent that we were in a position to consider potentially fruitful approaches to redesigning the Stat Lady tutor, with the goal of increasing the speed of use of the system. Additionally, the verbal protocol participants turned out to be „good“ subjects, who learned the material quickly and provided very few opportunities to study „problem“ areas in the way the curriculum was taught. A final factor that influenced our choice of strategies for improving learning efficiency is that large-scale studies (N > 100) indicate that average posttest scores are already satisfactorily high (Shute, 1995), and since the tutor is currently used strictly as a laboratory research tool, there is no outside pressure - from school districts, for instance - to try to raise posttest scores. Thus, with little to say on improving the curriculum itself, but some well-developed ideas regarding how to speed people through the current curriculum more quickly, we have decided to adopt the latter approach in our search for better learning efficiency.

2 Overview of the Stat Lady Tutor

Stat Lady is the name of a series of computerized learning environments teaching topics in introductory statistics (e.g., probability, descriptive statistics). This paper is proposing to modify one of the modules from descriptive statistics, namely "Data Organization and Plotting" (see Shute, 1995 for more detail than we are able to provide here). The design of all the Stat Lady programs reflects the theoretical postulates that learning is a constructive process, enhanced by experiential involvement with the subject matter that is situated in real-world examples and problems. The general goal of Stat Lady is thus to enhance the acquisition of statistical knowledge and skills by making the learning experience both memorable and meaningful. To accomplish this, each module was created to be a hands-on learning environment where learners actively engage in various on-line activities. Some of the features of the Stat Lady modules include: (a) a humorous and experiential interface; (b) an organized set of curriculum objectives; (c) a pool of whimsical, „real-world“ problem sets (Scenarios); and (d) a three-level feedback design, including both auditory and text feedback.

Stat Lady takes a mastery learning approach to teaching, and learning is self-paced. That is, a relevant concept or rule is presented, and then a problem is posed for students to solve in order to demonstrate comprehension (mastery) of each curriculum element. Further, the program is sensitive to (and thus addresses) errors that are recognized by the system's "bug" library. As mentioned, the particular Stat Lady curriculum relevant to this paper is descriptive statistics, assembled from the results of a cognitive task analysis performed by two subject matter experts (for validity and reliability), and consisting of 77 curriculum objectives. These were arrayed hierarchically, from simple to more complex concepts and skills, and then divided into five separate instruction Sections.

Students receive instruction and solve problems (Scenarios) within each Section, and Stat Lady provides specific feedback about each topic under investigation. More specifically, the program begins by identifying a particular topic for instruction from the list of curriculum objectives. After presenting the selected topic in broad strokes, Stat Lady quickly illustrates it with a real-life, humorous example to enhance memorability. Following instruction, learners have to apply the concept or skill in the solution of related problems. For problems that require data manipulation, they obtain their own unique data set from the Number Factory (an activity that's analogous to data collection), where they set their own parameters, such as the sample size and minimum and maximum values. If the learner successfully solves the problem on the first attempt, he or she is congratulated with positive auditory and textual feedback, and then moves on to the next problem or topic for instruction.[2] If a student is having difficulty (e.g., enters one or more incorrect responses), each response is followed by the increasingly explicit feedback (and encouragement), mentioned above. The system also allows learners to engage in elective "extracurricular" activities, such as viewing items in the on-line Dictionary and Formula Bank, playing around in the Number Factory, or using the Grab-a-Graph tool.

3 Proposed Changes

We have already mentioned two of the three sources of information from which we were able to draw implications about redesign of the tutor: verbal protocols and informal post-participation interviews. A third source of insight was the development of a production system model that reproduces the sort of adaptive changes in low-level interaction with the tutor (i.e., information-seeking behaviors) that we saw in the protocol subjects. The model was developed in ACT-R (Anderson & Lebiere, 1998), a cognitive theory of learning implemented in a simulation environment. Formal modeling of this sort forces one to make explicit the assumptions one has about goal structures and specific problem-solving paths. The assumption we made in this model was that our protocol subjects were satisfying the goal of figuring out which next action was appropriate at that point in the scenario (either by reading the directions off the screen or retrieving the next action from memory, based on past experience), and then executing that action. This seems like a reasonable interpretation of the behavior of the protocol subjects, largely because we know that they had, for the most

[2]There is an intelligent student modeling component (SMART - Shute, 1995), which tracks learner performance and can be relied upon to make decisions regarding which curriculum objectives need to be taught, when the student has mastered a particular topic, and so on. However, that capacity was disabled in the Gluck et al. (1998) study, as well as in this current proposal, so we do not describe it in detail.

part, already satisfied the other likely goal that someone using a computer-based tutor would have - which is to learn the curriculum. Thus, our underlying assumption is that someone learning from an ITS will be in the process of satisfying one of two goals at any point in the interaction: 1) learn the curriculum, or 2) figure out what action to take next (which may be done in service of learning a curriculum-related procedure).

The process of redesigning the tutor can be undertaken with respect to these goals. That is, we can make the assumption that any component of the tutor that is not critical for the successful completion of those two goals can be removed. This is essentially the approach taken here - to remove or modify a number of different features that were originally built into the system, and then examine the effect that removing them has on learning efficiency. These include changes to feedback sound files, the deletion of feedback text, the removal of the Number Factory, and the removal of references to problem Context. Next we'll describe each of these modifications in more detail.

3.1 Feedback Sound Files

Stat Lady offers both auditory and text feedback regarding performance on curriculum-related problems. Both the positive and negative feedback are designed such that a sound file plays first, then the text feedback appears. While the sound file is playing, the interface is "disabled," leaving the user incapacitated with respect to the tutor until it is done. The feedback sound files are selected randomly from a pool of options, and they range in length from about 250 ms to more than 8 s. It is the longer sound files that really become an issue. We observed repeatedly while running the protocol subjects that they would grow weary of waiting for the feedback sound files to finish playing, so that they could move on with the tutor. Since the hang-up really is with the length of the sound files, the modification to the tutor in this case is to replace them all with standard short (250 ms) sounds that represent positive and negative feedback. We measured the average length of the positive sound files to be 2.93 s and the negative sounds to be 2.29 s. Thus, each time a feedback sound plays, it would be 2-2.5 s faster than in the original version of the tutor. How much of an impact would this have on efficiency? We estimate that this simple manipulation alone will result in approximately a 7% decrease in tutor completion time.

3.2 Feedback Text

In the current design of the tutor, after the feedback sound file plays, feedback text appears on the screen. These are simple statements like, "Yes, that's correct!" or "Sorry, that is incorrect." If we take "learn the curriculum" and "figure out what to do next" as the two most likely goals while learning from Stat Lady (or just about any tutoring system, for that matter), then we see that the text we've been calling positive and negative feedback text achieves neither of these. There is no information about the curriculum or the next step. Due to the fact that subjects get the sound file feedback before the text, and therefore already know whether they were correct or incorrect, these messages are also redundant. Thus, they have very little information value, with respect to the likely goals of students using the tutor. It turns out that this interpretation is supported by the results of the protocol analysis, in that our subjects read a relatively low (and declining) proportion of positive feedback text as they solved problems in the scenarios. Figure 2 shows these data (Pos. Feedback) and also provides comparison data on reading behaviors for other

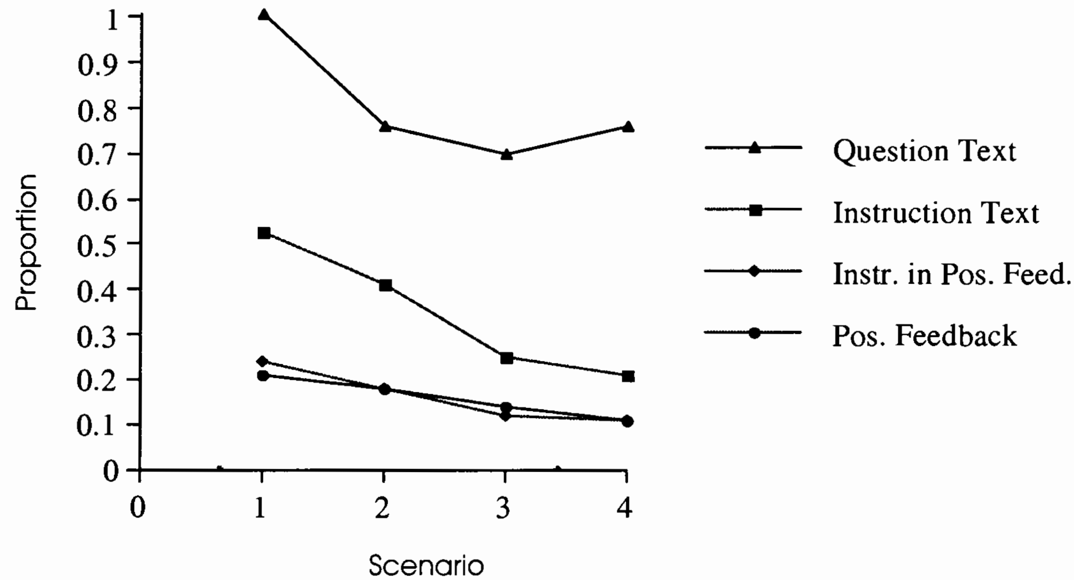

Fig. 2. Verbalization rates for four text types, averaged across scenarios in Sections 1 and 3 of the Stat Lady tutor. „Question Text" refers to the questions which appear in the scenarios to test learner ability on the curriculum objectives. „Instruction Text" contains information about what to do next and/or curriculum-relevant declarations. „Instr. in Pos. Feed." is instruction text that appears in the feedback window. „Pos. Feedback" is text that indicates whether the learner's response was correct or not.

types of text. Given the low information value of the positive feedback text, we are removing it from the tutor altogether. It seems like a theoretically and empirically justifiable change, but the impact on tutor time will be minimal. The estimate is about a 1/2% decrease in completion time as a result of this manipulation.

3.3 Number Factory

The Number Factory, which is designed as a metaphor for the process of collecting data, is an activity that is required at the beginning of each Scenario. The learner first reads the Context Statement, then goes off to the Number Factory to acquire data that will address the problem presented. Although it is an integral part of the original Stat Lady, it is not explicitly related to any curriculum objectives. There is no pretest or posttest assessment of the learner's skill in using the Number Factory, and there are not currently any data which bear on the issue of whether subjects come to associate „going to the Number Factory“ with actually conducting studies. Additionally, instructing students in the use of the Number Factory and visiting it at the beginning of every Scenario is a fairly time consuming process. It adds several minutes to the length of the instructional portion of Section 1 and requires approximately 45 s extra in every scenario. A natural alternative to visiting the Number Factory is to simply provide students with the data they need at the beginning of each scenario. Including the one-time savings associated with removing that portion of the instruction in Section 1, we estimate another 5% decrease in tutor completion time resulting from the removal of the Number Factory.

3.4 Situated Contexts

In reflecting on the various components of the tutor which add time, perhaps unnecessarily, to its completion, it becomes apparent that subjects spend an awful lot of time reading about "real-world" situations. As a result of the current excitement in educational circles regarding the benefits of situated learning (e.g., Lave, 1988), an effort was made in the original design of the Stat Lady tutor to situate the problem solving, both during instruction and during the Scenarios, in real-world contexts. References to context are pervasive throughout the tutor. They come up not only in the Context that appears at the beginning of each scenario, but also in the instruction sections, in questions during the scenario, and in feedback text. This is considerable additional reading time, and it remains an empirical question whether all this situated context actually has a positive impact on learning. A new version of Stat Lady that did not include the situated contexts in the instruction sections and the Scenarios could be used to test out that hypothesis. Regardless of the effect on learning, completion time is certainly going to drop as a result of this manipulation. A decrease in the neighborhood of 10% would not be unreasonable.

4 The Plan

We see the aforementioned changes to the Stat Lady tutor as developing along two dimensions. On the one dimension is the issue of Context: present or absent. This dimension is interesting with respect to learning efficiency, in that we predict faster progress through the tutor via elimination of the Context, and it has the added attraction of being relevant to one of the hottest debates in contemporary education research (see Anderson, Reder, & Simon, 1996, and the reply by Greeno, 1997). On the other dimension lies the proposed removal of various Features (feedback, Number Factory) which, quite frankly, are largely responsible for providing Stat Lady with a special personality all her own.

Both the Features and the Context manipulations serve the function of „structural“ redesign (Card, Moran, & Newell, 1983, Chapter 12) and merit evaluation. By crossing the two manipulations, we arrive at a 2x2 factorial design, with one factor being the presence or absence of Context, and the other being Features vs. No Features (more accurately, „fewer features"). Thus, this study involves comparisons of learning efficiency (outcome divided by time) across the following four conditions: 1) Stat Lady - Classic, 2) Stat Lady - NoContext, 3) Stat Lady - NoFeatures, 4) Stat Lady - NoContext *and* NoFeatures (super sleek new version). Table 1 summarizes our predicted time savings for each of these conditions.

Table 1. Predicted Time Savings (% Decrease) in the Four Conditions

Condition	Decrease in Completion Time
Stat Lady - Classic	Baseline
Stat Lady - NoContext	10.0%
Stat Lady - NoFeatures	12.5%
Stat Lady - NoContext *and* NoFeatures	22.5%

Outcome, the numerator in our learning efficiency computation, is perhaps better expressed as learning gain. It is the amount of improvement from the pretest to the posttest.

We mentioned briefly in the introductory paragraphs the issue of motivation. Ideally it is the case that any given learning experience is not only effective, but also enjoyable and rewarding. We acknowledge the possibility that either or both of the dimensions along which the tutor will change for this assessment could result in a significantly less enjoyable experience for the participants. Those data would be important, so we will include after the posttest an Affective Response Survey which queries the participants on the subjective experience of learning from the tutor.

5 Prospective Conclusions

The time savings predictions are based on performance data from previous studies. Since both manipulations essentially involve removing pieces of the tutor, we feel that it is safe to assume that their effect will be a decrease in tutor completion time, although the exact size of the effect remains to be seen. Few people would be surprised by this prediction. The real unknowns are what the impact of the modifications will be on learning outcome (posttest performance) and the subjective experience of the participants. The results are not at all obvious with respect to these two measures.

As one might anticipate, we do not expect these manipulations to have a negative impact on learning. If that were the case, it would be like shooting ourselves in the proverbial foot. Others might disagree, however, and quite possibly for good reason. Recent cognitive tutoring data, for instance, actually suggest some benefit is to be had in providing textual descriptions of problem-solving contexts similar to those in Stat Lady. One example of this is Koedinger and MacLaren's (1997) finding that students perform better solving a class of algebra word problems than they do on the analogous symbolic equations. Another relevant result is that greater emphasis on word problem solving in algebra instruction leads not only to greater posttest performance on word problem solving (no surprise there), but also to greater posttest performance on basic skills (Koedinger, Anderson, Hadley, & Mark, 1997). Perhaps it is the case, then, that posttest scores will be lower in a decontextualized, non-situated instructional environment. This is likely the prediction that the hardline situativists would make, but what if posttest scores do not go down? It will be interesting to consider the implications of such a result, if indeed that is what we find. We'll see what the data tell us.

It is worth noting that the pretest and posttest have contextualized problems. Thus, all students will be tested on contextualized problems and we will be seeing how the presence or absence of context during learning affects this transfer (to new cover stories for the Context group and to newly context-enriched problems for the NoContext group).

Finally, we consider the effects of our manipulations on subjective experience. Although there are a few exceptions (e.g., Simon, 1979), motivations and emotions and their interaction with learning are topics that have been largely ignored in cognitive psychology. In fact, Aubé (1997) recently commented on the need for better theories, and models based on those theories, of the interaction between emotional motivation and learning. If those models already existed, we might be able to generate some reasonable predictions of the effect our proposed modifications will have on subjective reactions to the new versions of Stat Lady. As it stands, however, we will again have to wait and see what sort of story unfolds in the data. We look forward to sharing that story at ITS '98.

References

1. Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought. Hillsdale, NJ: Erlbaum.
2. Anderson, J. R., Reder, L.M., & Simon, H. A. (1996). Situated learning and education. Educational Researcher, 25(4), 5-11.
3. Aubé, M. (1997). Toward computational models of motivation: A much needed foundation for social sciences and education. Journal of Artificial Intelligence in Education, 8(1), 43-75.
4. Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, NJ: Erlbaum.
5. Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data (revised edition). Cambridge, MA: MIT Press.
6. Gluck, K. A., Lovett, M. C., Anderson, J. R., & Shute, V. J. (1998). Learning about the learning environment: Adaptive behaviors and instruction. Unpublished manuscript.
7. Greeno, J. G. (1997). On claims that answer the wrong questions. Educational Researcher, 26(1), 5-17.
8. Koedinger, K. R., & Anderson, J. R. (in press). Illustrating principled design: The early evolution of a cognitive tutor for algebra symbolization. Interactive Learning Environments.
9. Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8, 30-43.
10. Koedinger, K.R., & MacLaren, B. A. (1997). Implicit strategies and errors in an improved model of early algebra problem solving. In Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.
11. Lave, J. (1988). Cognition in practice. Cambridge, UK: Cambridge University Press.
12. Shute, V. J. (1995). SMART: Student Modeling Approach for Responsive Tutoring. User Modeling and User-Adapted Interaction, 5, 1-44.
13. Shute, V. J., & Gluck, K. A. (1994). Stat Lady Descriptive Statistics Tutor: Data Organization and Plotting Module. [Unpublished computer program]. Brooks AFB, TX: Armstrong Laboratory.
14. Simon, H. A. (1979). Motivational and emotional controls of cognition. In Models of thought, Vol. 1. (pp. 29-38). New Haven: Yale University Press.