Effectiveness of Using an Intelligent Tutoring System to Train Users on Off-the-Shelf Software

Claver P. Hategekimana
Iowa State University, Human Computer Interaction and
Wenatchee Valley College, Teaching & Learning Center
USA
claver@iastate.edu or chategekimana@wvc.edu

Stephen Gilbert
Clearsighted, Inc. and Iowa State University, Human Computer Interaction
USA
stephen@clearsighted.net or gilbert@iastate.edu

Stephen Blessing
University of Tampa
USA
sblessing@ut.edu

Abstract: This study addresses the potential of using an intelligent tutoring system (ITS) to tutor on off-the-shelf (OTS) software. ITSs have been successfully used to tutor on a variety of learning domains, but there has been little research comparing ITS-based training on an OTS application with traditional software training approaches such as books or interactive software simulations. The work presented here includes procedures and results for training and evaluation using three methods: book-based, interactive simulation, and an ITS. We found that there were some associations between the training method and training experiences. Book-based training exhibited higher scores on both task performance and system usability perception, while better times were recorded for the simulation approach. Concept acquisition score was not found to significantly correlate with training method. We concluded that if an ITS is to be a tutor on OTS applications then further refinements are needed.

1. Introduction

Organizations have realized a great need for human resources development to cope with the gap between the skills of their personnel and increasing technological demands of jobs. Perhaps the biggest challenge within this issue is to find the most effective training programs. Following the work of Carr (2004), the use of computer applications for training is a stepping stone to competitive advantage when used effectively. Computer-based training is highly desirable, but the ability of trainees to transfer learning back to the job is frequently questioned (Baldwin, 1992). For the purpose of this study, we focused on how intelligent tutoring system (ITS) technology can be used to increase training outcomes of computer users that 1) produces a higher return on investment in terms of knowledge transfer, and 2) requires less training time and is therefore more cost-effective.

ITSs have been successfully used to tutor on a variety of domains such as mathematics, geometry, and economics (Aleven, et al. 2006; Arruarte,et al. 1997; Koedinger, 2004). The benefits of ITSs include personalized training, learning with real-world tasks, and multiple levels of help (Koedinger, et al. 1997; Arruarte, et al. 1997). All of these have potential to enhance training outcomes on OTS applications. The kind of ITS that we use here, a model-tracing tutor, relies on the principles of artificial intelligence and cognitive models embodied in Anderson's ACT-R theory and production rules (Murray, Blessing & Ainsworth, 2003; Anderson 1993).

Despite notable success stories, ITSs are not widely used in education and training programs outside high-school mathematics. There are two main reasons for this slow penetration: first, an ITS is expensive to develop; second, an ITS is generally evaluated according to its artificial intelligence criteria rather than with respect to educational effectiveness (Corbett & Anderson, 1997). If an ITS is to be a training model of choice for students and personnel using off-the-shelf software, validation of its effect on learning is critical.

Most ITS-based training developers have integrated the ITS with the training application itself, often creating the user interface from scratch to work seamlessly with the tutor. This current research stems from a larger project at Clearsighted, Inc. that attempts to link ITSs to OTS applications to create better training environments than currently available. Using this technique, students using the system interact with the actual software on which they are being trained, not a simulation or some other facsimile.

All three of the learning conditions are self-paced, which Simon and Werner (1996), and Anderson (1995) argue that computer users prefer for training, especially for computer-related training. Self-paced instruction provides computer users with greater opportunity for practice than the instructor-led training approach. Simon and Werner indicated that lectures appear to be stronger in providing trainees with understanding, although generally without the opportunity to practice as provided by other techniques, e.g., computer users can often figure things out for themselves.

Below we describe a December 2007 experiment involving three different learning conditions: 1) book-based, 2) interactive simulation movies, and 3) model-tracing ITS. We chose book-based instruction because it is a widely-used method for learning computer applications. We also chose interactive simulation movies built with Adobe Captivate 2 not only because they are currently a cutting edge technology that simplifies the task of creating training demos and simulations, but also because simulations permit the attainment of learning goals which are beyond traditional methods (Hooper & Thomas, 1991). However, these training media often offer relatively passive learning experiences which are not personalized.

2. Methods

2.1 Participants

Participants in the experiment were undergraduate students from a public school in the Midwest. Potential participants were invited to register online and were filtered based on previous Adobe Photoshop experience and major (in particular, excluding computer science and graphic arts majors). Thirty five dollars and a chance to win one of four iPod Nanos were used as incentives. The study required a commitment to attend both a 120-minute training session and a 90-minute testing session one week later.

The sample consisted of 75 participants of whom five did not disclose their demographic information. We identified 39 male and 31 female participants. Sixty nine were in the 18-24-age group and one participant was in the 25-35-age group. All participants reported that they use computers everyday. Most of them (n=50) described themselves as average computer users, though there were19 advanced users and 1 novice computer user. None of them had experience with Paint.NET or was familiar with an intelligent tutoring system. Their levels of education ranged from freshman (16), sophomore (21), junior (20), and senior (11). Participants were from a wide variety of majors including business (18), engineering (15), educations, and psychology (5).

2.2 Materials and Procedure

The target application for training was Paint.NET, an open source application with basic features similar to Adobe Photoshop. Training and testing experiment sessions were administered in a spacious and modern campus classroom equipped with 40-networked computers. To allow for participants' schedules and the room capacity, there were four identical training sessions and four identical testing sessions that followed. Participants were randomly assigned to one of the three training conditions. After completing training, they returned at the same time one week later to complete three test tasks in Paint.NET without support. Training sessions started with a welcome message. Participants were given an overview of what the experiment was about with emphasis on the sequence of the events during the training and the testing sessions. Other information the facilitator provided included incentives, confidentiality, and dealing with technical issues (e.g., what to do if the application freezes). They were also advised that they could answer or leave out any question they feel uncomfortable with or withdraw at any time without penalty.

Next, all participants watched a video clip that introduced them to the basic features of Paint.NET (headphones were provided) and filled out a questionnaire for demographic information. They then completed 12 training tasks with scaffolding based on their assigned conditions (book, simulation, ITS). Time on each training task was recorded. Additionally, they completed a 10-item system usability questionnaire developed based on the scale of Brooke (1996). Finally, they completed a 3-item attitudinal questionnaire about their experience:

- What are three things you liked about the training you received?
- What are three things you did not like about the training you received?
- Did the training you received help you use Paint.NET better than you would have without it? If yes, how? Otherwise, what do you see to be its weaknesses?

After one week, participants returned to the same computer lab and completed a questionnaire designed to test conceptual knowledge of image manipulation as well as three performance tasks in Paint.NET. Testing sessions ended by giving participants $35 compensation for their time and announcing the winners of the iPods. To ensure fairness, a computer application randomly picked a winning number among the 4-digit student identification numbers used throughout the study.

Because the research question focused on the effectiveness of three training approaches, identical content was used in all three media as much as possible (see strengths and limitations in Table 1). At the same time, the natural affordances of each medium led to different characteristics in the training experience. It is typically easier to flip back and forth among pages in a book, for example, than to refer back to previously viewed video footage. All three training methods used a learning-by-doing approach. Tasks featured a brief motivating scenario, "*What's the Point"* and a high-level goal and specific requirements, as well as background information labeled "*Concepts.*" The three methods differed primarily in how they presented the steps for reaching the goal. The results were analyzed to establish the effectiveness of each training method using four measures:

- Total training time
- Concept acquisition score
- Task performance score, and
- System usability perception

Table 1: Comparison of strengths and limitations of book, simulation and intelligent tutoring system

Opportunity for strength or weakness	Book	Simulation	ITS
• Interaction with software you are learning (Paint.NET)	High	Low*	High
• Learn-by-doing (use a real world exercise for learner to practice)	High	Medium*	High
• Self paced instruction (adjust to learner's speed)	High	Medium**	High
• Detailed step-by-step instruction on how to complete the task	High	None	None
• Demonstration by using example (vicarious learning)	None	High***	None
• Help on demand (built-in Paint.NET help was disabled)	None	Low	High
• Just in time help (automated timely feedback)	None	High	High
• Multiple paths to the solution	None	None	High
• Alternate the sequence of tasks	High	None	High
• Revisit previously viewed material	High	None	None

** Learner interacts with screen shots of Paint.NET rather than the actual Paint.NET software.*
*** Learner can not speed up or slow down the transition time between slides.*
**** Learner watches a demo that introduces new skills needed to complete the training tasks.*

3. Results and Discussion

Data were analyzed by using SAS general linear models procedures (SAS Institute, 2003). There was one independent variable (training method) and four dependant variables (total training time in minutes, concept acquisition score, task performance score and system usability perception score). Results indicated that training condition has the potential to significantly influence training experience. Our expectation in the study was that the ITS group would learn the material better, in terms of total training time and post-test measures, than the other groups. What follows however, are the contrary results from our December 2007 experiment (see descriptive statistics in Table 2).

Table 2: Comparison of participant scores on measures of training mode success

	n	Min	Max	Mean
Total training time in minutes ($F(2,52)=3.59$, $p=.04$)				
Book	17	37.12	71.42	56.86a
Simulation	24	30.83	75.80	46.85b
ITS	14	31.33	80.70	57.60a
Concept acquisition score ($F=(2,72)=1.459$, $p=0.239$)				
Book	24	4.00	9.00	6.92a
Simulation	26	1.00	10.00	6.12a
ITS	25	2.00	9.00	6.12a
Task performance score ($F(2,64)=3.859$, $p=0.026$)				
Book	18	11.00	40.00	25.00a
Simulation	25	4.50	35.00	17.70b
ITS	23	6.00	34.50	22.56a
Usability perception score ($F(2,71)=7.611$, $p=0.001$)				
Book	24	45.00	76.50	60.28a
Simulation	26	40.50	72.00	59.89a
ITS	24	33.75	69.75	50.91b

Means followed by the same letter within column are not significantly different (p>0.05)

On the task performance score the book group (*M=25.00*) and the ITS (*M=22.56*) were significantly different than the score of simulation group (*M=17.70*). The scores of book and ITS however, were marginally significantly different. This finding was consistent with the results of a December 2006 pilot study with seven participants where the means of the book and the simulation groups on both concept and task performance scores were not significantly different at the outset.

Likewise, the overall system usability perception rating for the book group (*M=60.28*) was significantly better than the rating for the ITS (*M=50.91*); but the ratings of book and simulation were not statistically significantly different. This was not the case during a December 2006 pilot study where simulation group significantly reported their system as more usable than the book group. Due to the small sample size (n=7); however, it was hard to draw strong conclusions from the pilot study.

Despite this, the present data showed that the simulation group (*M=46.85*) used less total training time (less time being desirable) than both the book (*M=56.86*) and the ITS (*M=57.60*), a significant difference while the total training time for book and ITS were marginally significantly different.

On the open-ended questionnaire, almost all participants in all conditions expressed that the training they received help them use Paint.NET better than they would have without it. The systems showed them where to find the features of the software, introduced them to new concepts and provided hands experience that taught something new. Only two out of 24 participants in the book condition and two out of 26 participants in the simulation indicated that the training modes they received were not helpful. For example, unsatisfied book participants felt that they would have had less frustration if there were offered a different type of training, otherwise they would have preferred to experiment on their own rather than following every step outlined in the book. At the same time frustrated simulation participants expressed that they would have preferred freedom to struggle while discovering how Paint.NET works rather than being forced to one way of doing things when there could be other alternatives. Probably this was a matter of learning preferences that was not well matched with the assigned training conditions.

While participants with book-based instruction were told exactly what to do and enjoyed skimming and sometimes skipping background instructions, they also reported frustration completing tasks. The book was criticized because it was not interactive, did not provide sufficient help, and was much like a long and arduous lab assignment. Participants ended up skipping some steps, which somewhat hurt their task performance even though they performed equally well as the ITS group and better than the simulation. Alternating back and forth between physical pages and the computer screen was also seen as a downside. However, none of these criticisms was repeated at an alarming rate.

The majority of simulation participants indicated that their system was clear in giving instructions. Unlike book-based instructions, clicking through the simulations and seeing the steps that participants would need to do to succeed (vicarious learning) was an advantage. They also liked just-in-time help, a pop-up window that provided instruction in case of error. Their system controlled the learning environment rather than the learners because the option to proceed to the next step depended on the success of current step. This guaranteed successful completion of the task, but it minimized the amount of creativity. About 95% of the participants felt like simulations: 1) did most of the work, 2) did not teach the concepts well enough, 3) did not support multiple paths through the software, and 4) presented tasks and steps in a fixed linearly fashion. About a half of the participants indicated their system: 1) did not offer the same level of flexibility to refer back to previously presented materials as the printed material allowed users to skim instructions, and 2) did not provide option to speed up or slow down the transition time between slides to accommodated speed of faster or slower readers.

Similarly, the ITS group enjoyed a high level of creativity because their system supported multiple paths to the solution. The most appreciated feature of ITS was the hint function. Unlike the book and the simulation, ITS participants could ask for a hint which presented different levels of guidance. There was also just-in-time (JIT) help that displayed a message when a mistake was made. Participants could alternate the sequence of tasks, but a sequential order was encouraged. However, almost all participants expressed that the ITS did not offer an exhaustive list of paths to the solution. It forced users to successfully complete a given subgoal in order to proceed to the next subgoal. Some reported that it was too sensitive in displaying error messages. Other complaints were technical issues including frequent application freeze.

The interesting results come from the ITS's ability to track learners' actions. It was observed that participants visited all additional content (What's the Point and Concepts information) for the first few tasks. As they learned how the system worked, different participants selected different paths. Some participants relied only on What's the Point direction that described the goal of the exercise and ignored the Concepts link. Also, many participants frequently asked for hints as they did not have step-by-step instructions to guide them. Fortunately, the different levels of requested help not only provided guidance leading to the solution, but also offered enrichment with conceptual explanation. Arguably, the help provided by ITS was comparable to the step-by-step instructions supported by book. The results exemplified that currently supported multiple paths was a matching approach to step-by-step instruction supported by book. A higher number of paths to the solution would have potentially resulted in a more favorable result on task performance than the book and simulation. It would also have minimized the frequency of requesting help which had direct effect on the time spent on the training tasks. Minimizing technical issues (e.g., application freeze) and better curriculum presentation (e.g., unconditional transition to the next task and offering the option to follow step-by-step instruction) could have minimized frustration, which had direct effect on system usability perception and concept acquisition scores.

It is hoped that the information from the current study will make it possible for ITS designers to improve the technical component of the system, presentation of the curriculum, and system usability. Refinements should lead to improved measures on educational system effectiveness; otherwise the alternative systems we examined seem to be the best options.

4. References

Aleven, V., McLaren, B., Roll, I. & Koedinger, K.R. (2006). Toward meta-cognitive Tutoring: A Model of Help-Seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education*, 16, 101-130.

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

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

Arruarte, A., Fernandez-Castro, I. & Greer, J.E. (1997). The IRIS Shell:"How to Build ITSs from Pedagogical and Design Requisites. *International Journal of Artificial Intelligence in Education*, 8, 341-381.

Baldwin, T.T. (1992). Effects of alternative modeling strategies on outcomes of interpersonal skills training. *Journal of Applied Psychology*, 77, 147-154.

Brooke, J. (1996). A “quick and dirty” usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester & A . L. McClelland (eds.). *Usability Evaluation in Industry*. London: Taylor and Francis.

Carr, N.G. (2004). *Does IT matter?: Information technology and the corrosion of competitive advantage*. Boston, MA; Harvard Business School Press.

Corbett, A.T. & Anderson, J.R. (1997). Intelligent tutoring systems. *Handbook of Human Computer Interaction*. Second Ed. In M. Helander, T.K. Landauer, P. Prabhu (Eds), Elsevier Science B.V, Chapter 37.

Hooper, E.J. & R.A. Thomas, R.A. (1991). Simulations: An opportunity we are missing. *Journal of Research on Computing Education*, 23(4), 495-513.

Koedinger, K.R., Aleven, V., Heffern, McLaren, B. & Hockenberry (2004). Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration. In the *Proceedings of the Seventh International Conference of Intelligent Tutoring Systems*, Maceio, Brazil.

Koedinger, K.R., Anderson, J.R. (1997). Intelligent Tutoring Goes to School in the Big City. *International Journal of Artificial Intelligence in Education, 8*, 30-43.

Murray, T., Blessing, S. & Ainsworth, S. (2003). *Authoring Tools for Advanced Technology Learning Environments*. Netherlands.

SAS Institute (2003). *SAS User’s Guide. Statistics, Version 9.1*, Cary, NC. SAS Institute Inc.

Simon, S.S. & Werner, M.J. (1996). Computer Training Through Behavior Modeling, Self-Paced, and Instructional Approaches: A Field Experiment. *Journal of Applied Psychology*. (81) 6, 648-569.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. OII-0548754. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors are also thankful to Wenatchee Valley College for providing the funding needed to attend and present this work to the SITE 2008 Conference.