DOCUMENT RESUME

ED 357 065 TM 019 796

AUTHOR	Orey, Michael
TITLE	Three Years of Intelligent Tutoring Evaluation: A Summary of Findings.
PUB DATE	Apr 93
NOTE	12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 11-16, 1993).
PUB TYPE	Information Analyses (070) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Armed Forces; *Computer Assisted Instruction; Computer Software Development; Computer Software Evaluation; Evaluation Research; Feedback; *Instructional Effectiveness; *Learning Processes; Problem Solving; Teacher Role; *Technological Advancement; *Tutoring
IDENTIFIERS	*Intelligent Tutoring Systems; MALM Tutoring System; MSRT Tutor; POSIT Tutoring System

ABSTRACT

Over the past 3 years, a variety of studies in intelligent tutoring system (ITS) effectiveness have been conducted. A summary is provided of the research into the use of POSIT, MALM, and the Mobile Subscriber Remote-Telephone Terminal (MSRT) Tutor. POSIT is an ITS for the tutoring of whole-number subtraction. It assumes that the learning of a cognitive skill builds from declarative knowledge. MALM, which makes use of some of the latest computer-based instructional technologies, was designed so that problem solving in an Army communications network can be explored in a computer environment. The MSRT Tutor teaches operating procedures for one of the Army's mobile telephones. Studies conducted with each of these ITS systems suggest that an ITS has the potential for being effective for instruction, especially with procedural knowledge. An ITS may be used as a change device to change the role of the classroom teacher from one that is largely disciplinary to one that is more facilitative. Help and error feedback seem to be the most useful features of an ITS. The use of off-the-shelf development packages is a great boost to the use of ITSs, since hours of design and development can be cut short. (SLD)

Three Years of Intelligent Tutoring Evaluation: A Summary of Findings

by

Michael Orey

The University of Georgia

Department of Instructional Technology

607 Aderhold Hall

Athens, Ga 30602-7144

(706) 542-3810

Paper presented at the annual meeting of AERA, April, 1993, Atlanta, Florida.

Running Head: THREE YEARS

Three Years of Intelligent Tutoring Evaluation: A Summary of Findings

Over the past three years we have conducted a variety of studies on ITS effectiveness (Orey, 1991: Orey, et al., 1992b; Orey, Trent & Young, 1992). Each of these systems have employed a model tracing approach to its design, so these evaluations will also reference other systems that take this approach (see, Anderson, Boyle, & Reiser, 1985). The evaluations have focussed on overall effectiveness (Orey, 1991), effectiveness of coaching or help (Orey, et al., 1992b), the effectiveness of error feedback (Corbett & Anderson, 1991; Orey , et al;., 1992b), and the streamlining of the ITS development process (Orey, Trent & Young, 1992). Two areas examine the most "intelligent" characteristics of an ITS - the impact of error feed back and the impact of coaching or help. An ITS is designed so that it will dynamically adapt its instruction to an individual learner. The way that they do this is by providing help or coaching that is relevant to that learner in the particular situation that the learner currently finds herself. In addition, the system has some form of error diagnosis, and provides feedback on errors as they occur. These aspects of ITS have been examined in the context of three different tutors: POSIT, MALM, and the MRST Tutor. I begin this paper with a brief description of each of these systems.

Description of POSIT

POSIT is an ITS for the tutoring of whole-number subtraction. The design of this system (see, Orey & Burton, 1989/90, for a detailed description of the design) is based somewhat on Anderson's (1987) ACT* theory of learning. Within this theoretical orientation, it is assumed that the learning of a cognitive skill builds form declarative knowledge. In terms of subtraction, the algorithm to be learned is made of a set of declarative information. During the development of this skill, this knowledge becomes consolidated into the specific skill (in this case, proficiency with a subtraction algorithm develops). Declarative knowledge is provided to the learner in a text form when errors occur or when the learner asks for help. For example, this is a typical message after making the Error of Omission - Decrement (#1 and #2 are variables that are bound during the execution of the program) "*You have to complete the borrow into the ones place by taking away one of the tens. So, the correct value for this area is #1 You typed #2 Please enter the correct value.*" Although this is a very brief description of POSIT, it should suffice for the purposes of this paper. POSIT was developed in Lisp on a Macintosh with 2 megabytes of RAM.

Description of MALM

The architecture of MALM was based on a problems solving perspective and it makes use of some of the latest computer-based instructional technologies: ITS, leaning environments, and hypermedia. These techniques are combined so that problem solving in an Army communications network (MSE is its name) context can be explored in the somewhat safe and inexpensive environment of an IBM PC computer. MALM was designed for high bandwidth diagnosis using a model tracing technique (Van Lehn, 1988). It was implemented in a hypermedia environment that allowed the learners to explore the environment, much like that in a learning environment. The knowledge base representation combines a frame based representation and a rule based implementation. Problems are generated using an automated parsing process and instruction is provided when errors and help occur. These aspects of MALM were added (for a more complete description of MALM, please refer to Orey, et al., (1992a) to the already existing system that was developed by Galaxy Corporation (Coonan, Johnson, Norton, & Sanders, 1990). For the first two experiments that involved MALM, we used human tutors, ITS and CAI approaches. What this means is that we altered MALM so that it would behave like a CAI (called here the Linear Advice (LA))program and we altered it so that it could be used by human tutors. The result is that all three groups has access to a well engineered computer program. This would cut down on the effect of the MALM ITS in comparison to the other groups. If we take Kulik and Kulik (1987) at face value, then the effect size estimate for the LA version of MALM would actually be slightly higher than 0.31. In addition, well qualified tutors were not available for the experiments, so the engineers of MALM were used to perform tutoring tasks. The effect would be to lower the 2.0 effect size. Therefore, the comparisons between groups would be harder to detect differences because they are closer to same effect size. MALM was developed using the C programming language. It was designed to run on an MS-DOS computer with and EGA monitor and a minimum of 640 kilobytes of RAM.

Description of the MSRT Tutor

The context was operations procedures for one of the Army's mobile telephone (the Mobile Subscriber Remote-telephone Terminal or MSRT). We examined two different development environments--an off the shelf hypermedia tool and the C programming language. Two programmers worked on this project. One used IBM's LinkWay® to develop a hypermedia-based ITS. The other programmer used the collection of C routines that had been used in MALM. The idea was that considerable time savings could be achieved if the developed of the new tutor (the MSRT tutor) used the existing C routines (libraries). The original tutor, MALM, essentially was a hypermedia based ITS. Its primary functions were to provide appropriate advice throughout the problem solving process and to provide corrective feedback. The second programmer chose to use a hypermedia tool and build the tutor from scratch. The constraint for both systems was that the system must run on an MS-DOS machine in EGA mode and use 640 K or less or RAM. LinkWay® is a hypermedia tool that meets those requirements. From here on out, we will refer to the LinkWay version as MSRT-L and the version based on MALM and written in C as MSRT-C.

The Overall Effectiveness

It would be ludicrous to try to generalize from a single evaluation to all instances of ITS, so I will describe the results of the POSIT evaluation in the light of the results of similar studies. Legree and Gillis (1991) have performed a synthesis of large scale evaluations of ITS technology in a variety of fields. Most cf these systems focus on the acquisition of procedural skills and related problem solving application. The populations vary from children to college students to military trainees. The general conclusion of this analysis was that the effect size of ITS as an instructional strategy relative to large group lecturing, is about 1.0. This estimate is based on only three large scale studies. Therefore, the 1.0 effect size can be considered preliminary. They also recommend that methodology for comparisons ought to include three treatments: a large lecture type class condition, a tutorial condition, and the ITS condition. POSIT was used in just such a study.

POSIT was used in such a study. However, of interest in this study was the impact of implementing the same instructional strategy in different ways. In one condition, a mastery learning condition was implemented in a large group condition, a small group with a tutor condition (3 or fewer students), and the ITS condition. Results were not conclusive. The primary dependent measure was time to achieve mastery and the associated probability was 0.086. Possible limitations to these results were that the volunteer tutors were engaged for a fixed time and they terminated the experiment at the end of this time. Unfortunately, the study was not yet complete. While all of the POSIT group participants had achieved mastery, only 8 of 12 in the large group and 11 of 12 in the small group tutoring conditions had achieved mastery. The result was that there were only 8 participants in each group and that small of a sample makes it difficult to determine differences. However, because of the exploratory nature of this study and because the probability was close to our chosen alpha level, we decided to go ahead and calculate effect sizes. Relative to the large group mastery learning condition, the ITS group had an effect size of 0.74. Because this is relative to a mastery learning treatment, I would estimate (although not reliably because of the lack of significance) that this figure would be close to the 1.0 predicted by Legree and Gillis (1991).

Perhaps the aspect of the POSIT evaluation that was most interesting was the kinds of interactions that occurred between teacher or tutors and the students in each of the treatments. It was found from video tapes of two different sessions in each treatment that the majority of interactions between the teacher and students in the large group condition dealt with the correction of inappropriate behavior. The majority of interactions between the tutors and the ITS lab assistant and their students were more of an academic assistance nature. For most teachers, this would be THE most significant outcome. The role of the teacher changes from one of behavioral manager to collaborator in learning. This observational result is similar to that found by Schofield, Evans-Rhodes, & Huber, (1989) who did an ethnography of a school that implemented a geometry tutor.

The Intelligent Aspects of ITS

As stated above, the two attributes of ITS that exhibit the most intelligence are the help function and the error feedback function (especially in model tracing tutors). While we have done some work on the analysis of error feedback, most of our work has focussed on help (or advice or coaching). POSIT was evaluated in terms of how well it performed diagnosis (Orey & Burton, 1990). While POSIT was found to be quite effective at diagnosis (76% correct versus about 50% correct with other systems), the impact of this diagnosis was not directly examined. However, Corbett and Anderson compared the effects of a variety of different approaches to feedback, but found only that feedback was better than not having feedback.

The effects of the advice or coaching function were examined in a series of studies that we conducted using MALM (Orey, et al., 1992a; Orey, et al., 1992b). Essentially, this series of studies used three groups. One group learned army telecommunications via the ITS. Another group used a simulation that used the same simulation as the ITS but with the advice and error checking turned off. The last group used a simulation version that provided a screen listing all the steps in the procedure. It was left to the learners to determine what they had done and what they had yet to do . Also, the error detection system simply indicated that an error had occurred, but did not elaborate on what the error was. In the initial two studies, the only interesting result was the ITS group tended to use the advice function much more frequently. The third study was designed to examine this phenomena in greater detail.

Employing a methodology borrowed from social interaction research (Allison & Liker, 1982), we constructed a study to examine the impact of advice on the performance of the learners in the learning environment (Orey, et al., 1992b). Essentially, the procedure involves comparing the learner's behavior following advice (the computer's behavior) versus the learner's behavior following an event where the computer does not provide advice (or any other instructional behavior). In addition, the behavior of the learners in the ITS group can be compared to the behaviors in the group who only got the "screen full of steps" form of (called the Linear Advice (LA) group). Results indicated that learners's tended to perform activities that were directed to solving the problem for at least two behaviors following advice in the ITS group and only one step in the LA group. However, when comparing the two groups, there were no meaningful statistical differences. This collection of studies perhaps ought to give us pause. However, an easy explanation is that the treatment effects may be important, but the poser of the difference is not so great as to be able to measure the difference on the testing instrumentation available in the social sciences.

Development Streamlining

Perhaps the most important criticism that has been leveled against the field of ITS has been in the area of development time. Estimates range from 100 to 500 hours of development time is needed for every one hour of ITS instruction that is developed. This development time is inordinate and impossibly expensive for many applications. While there are some who are exploring the streamlining of development for ITS (Towne, 1991), little is now available. This systems described above took 200 hours per hour (POSIT, written in Lisp) and 200 hours per hour (MSLM, written in C). In the more general area of computing visual programming environments (like HyperCard, ToolBook, and LinkWay Live!) are streamlining the development time considerable, A valid question to consider is whether an ITS could be developed in this kind of environment and if so, is it as effective in delivering instruction. To test this idea, Orey, Trent , and Young (1992) set about developing the exact same system in two different environments. One system was developed by stripping out of MALM its content and putting in the MSRT content. The other system was built from scratch in the Link-Way Live! environment. Results of this experiment indicated that it took 2.4 times longer to develop the system in the MALM architecture than it did to develop it in LinkWay Live! Further, the system was given to soldiers who need to know how to operate the

MSRT. Some were asked to use the MALM version and other the LinkWay Live! version. The performance measures were about the same, although the LinkWay Live! group made fewer errors than the MALM group. However, the most interesting result and method was that we asked the soldiers to examine the "other" system. All but one person preferred the LinkWay Live! version. We anticipated that after spending a couple of hours with a system that people would build an allegiance to that system, but clearly only one person was unwilling to break that allegiance. Therefore, the LinkWay Live! version was preferred by the users, it may be more effective and it took less than half of the time to build (it is estimated that it took about 80 hours to develop each hour of instruction).

Conclusions

There are several points to review here. It seems that ITS has the potential of being effective for instruction, especially with procedural knowledge. In addition, it appears that ITS may be used to as a change device to change the role of the classroom teacher from one that is large disciplinary in nature to one that is much more facilitative. Help and error feedback seem to be useful features and tend to be the most intelligent aspects of the ITS system. Finally, the use of off-the-shelf development packages seems to be most powerful of conclusions. One of the greatest limitations of ITS has been that it takes many hundreds of hours of design and development for one hour of instruction. Also, ITS development required a PhD in computer science and artificial intelligence. This can be resolved through the use of off-the-shelf development environments such as hyperCard and LinkWay Live! It seems as though there are no barriers to further ITS development.

References

Allison, P.D., & Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin,* 91(2), 393-403.

Anderson, J.R, (1987). Skill acquisition: Compilation of week-method problem solutions. *Psychological Review,* 94(2), 192-210.

Anderson, J. R. , Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science* 228, 456-462.

Coonan, T. A., Johnson, W. B., Norton, J.E., & Sanders, M.G. (1990, February). *A hypermedia approach to technical training for the electronic information delivery system.* Paper presented at the meeting of the Society for Applied Learning Technology, Orlando, FL.

Corbett, A. T., & Anderson, J.R. , (1991, April). *Feedback control and learning to program with the CMU lisp tutor.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kulik, J.A. & Kulik, C.C. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology.* 12, 222-230.

Legree, P.J., & Gillis, P.D. (1991). Product effectiveness evaluation criteria for intelligent tutoring systems. *Journal of Computer-Based Instruction,* 18(2), 57-62.

Orey, M.A. (1991). External evaluation of POSIT. In D.W. Dalton (Ed.), *Proceedings of the 33rd International ADCIS Conference,* (pp. 3650373). St. Louis, MO.

Orey, M.A., & Burton, J.K. (1989/90). POSIT: Process oriented subtraction interface for subtraction. *Journal of Artificial Intelligence in Education,* 1(2), 77-104.

Orey, M. A., Park, J.S., Chanlin, L. J., Jih, H., Gillis, P. D., Legree, P. J., & Sanders, M. G. (1992a). High bandwidth diagnosis within the framework of a microcomputer-based intelligent tutoring system. Journal of Artificial Intelligence in Education, 3(1), 63-80.

Orey, M.A., Park, J.S., Chanlin, L.J., Gillis, P., & Legree, P. (1992b). *Does ITS Help, Help? and, Is an ITS Error Diagnosis a Remedy?*. In H. Troutner (Ed.), *Proceedings of the Thirty-fourth Association for the Development of Computer-based Instructional Systems*, (pp. 517-530). Norfolk, VA: ADCIS.

Orey, M.A., Trent, A., & Young, J. (1992). *Development efficiency and effectiveness of alternative platforms for intelligent tutoring for the mobile subscriber radio-telephone terminal* (Contract No. DAAL03-91-C-0034, TCN No. 92-287), Washington, D.C.: Scientific Services Program, Army Research Institute.

Schofield J. W., Evans-Rhodes, D. & Huber, B. R. (1989).Artificial intelligence in the classroom: The impact of a computer-based tutor on teachers and students (O.N.R. Technical Report No. 3). Arlington, VA: Office of Naval Research.

Towne, D. (1991). *Instruction in a simulation environment: Opportunities and issues*. Paper presented at the annual meeting of the Intelligent Computer Aided Training conference. Houston, TX.

VanLehn, K. (1988). Student Modeling. In Polson, M.C. & Richardson J.J. (Eds.). *Foundations of intelligent tutoring systems*. (pp. 55-78) Hillsdale, NJ: Lawrence Erlbaum Associates Inc.