

MACH III: Past and Future Approaches to Intelligent Tutoring'

Sylvia Acchione-Noel

U.S. Army TRADOC Analysis Command -
White Sands Missile Range

Joseph Psotka

U. S. Army Research Institute

Abstract

In 1986, the U.S. Army Research Institute created an intelligent tutoring system as a proof-of-concept for artificial intelligence applications in Army training. The Maintenance Aid Computer HAWK Intelligent Institutional Instructor (MACH III) taught student mechanics to maintain and troubleshoot the AN/MPQ-57 High Power Illuminator Radar (HPIR) of the HAWK Air Defense Missile System. In 1989, TRADOC Analysis Command compared the effectiveness of MACH III to traditional paper-based troubleshooting drills. For the study, all students received lecture and hands-on training as usual. However, during troubleshooting drills, students traced faults using either MACH III or the traditional paper-based method. Class records showed that the MACH III group completed significantly more troubleshooting tasks and progressed through tasks of greater difficulty than the paper-based group. Upon completion of training, students took written, practical, and oral essay tests. Mean test scores showed that students performed similarly regardless of the drill method used. However, significantly different standard deviations showed that the MACH III group performed more consistently than the paper-based group. Furthermore, significantly different time measures showed that the MACH III group reached faster troubleshooting solutions on the actual radar transmitter than the paper-based group. We will present the study results and discuss how updating the design of MACH III can include desktop computing in a virtual environment.

Introduction

In 1986, the U. S. Army Research Institute created an intelligent tutoring system (ITS) as a proof-of-concept for artificial intelligence applications in Army training. Called the Maintenance Aid Computer HAWK Intelligent Institutional Instructor (MACH III), it supported transmitter and receiver instruction in a radar maintenance course. The MACH III taught student mechanics to maintain and troubleshoot the AN/MPQ-57 High Power Illuminator Radar (HPIR) of the HAWK Air Defense Missile System.

The design of MACH III continued the philosophy behind STEAMER (Hollan, Hutchins, and Weitzman, 1984; Psotka, Massey, and Mutter, 1988). MACH III designers focused on user mental models; they emphasized conceptual rather than physical fidelity; and they aimed to construct generic tools (e.g., the conduit program) as much as possible. Also, in the STEAMER tradition, MACH III provided students with a graphical interface to interact with an inspectable simulation and a troubleshooting expert system. Just as STEAMER graphics displayed the movement of steam through pipes, MACH III graphics displayed the movement of electric current through radar components.

Students interacted with MACH III using its keyboard, mouse, and two monitors. Interaction occurred through three different modes: magic mode, real-life mode, and demonstration mode. The magic mode permitted the students to test and replace a variety of radar components through direct interaction with the model-based

simulation and view the results. Its "magic" nature enabled students to explore components that they ordinarily could not explore on the radar itself. In the real-life mode, interaction with the model-based simulation was mediated through the troubleshooting expert system. Students could test and replace components only as they would on the radar. However, they could receive one of three types of feedback: advice, critique, or both advice and critique. Also, in the demonstration mode, students could request the troubleshooting expert system to perform the next appropriate step(s) of a task.

MACH III was designed to help student mechanics develop the appropriate mental models for troubleshooting a radar. If successful, it would enable students to conceptualize radar signal loops-within-loops and to apply this knowledge to obtain faster, more efficient solutions. In 1989, TRADOC Analysis Command compared the effectiveness of MACH III to traditional paper-based troubleshooting drills (Acchione-Noel, 1991a; Acchione-Noel, 1991b; Acchione-Noel, Saia, Williams, and Sarli, 1990). The first half of this paper reports that quantitative evaluation of MACH III in a real-world setting under controlled conditions.

Method

The evaluation compared the training effectiveness of two courses. The traditional course contained lectures, training on the radar itself, and paper-based troubleshooting drills. In the paper-based drills, students traced symptoms through the manuals and schematics. The other course also contained lectures and training on the radar, but students performed troubleshooting drills with MACH III. In this case, students traced a symptom and tested components simulated by MACH III software. Manuals served as references. Lectures and radar training remained identical for the two groups. Therefore, any differences in performance could be attributed to the supplemental instruction. The evaluation focused on which method provided effective and efficient training--troubleshooting on paper or troubleshooting on MACH III.

Twenty-nine students with American citizenship and no previous radar maintenance skills participated in the data collection between December 1989 and May 1990. Previous exam scores and other demographic data were used to achieve stratified random assignment of students to the two courses.

Instructors of the MACH III and paper-based groups recorded each student's progress through a troubleshooting task list created for the study. The troubleshooting task list contained a lengthy list of easy, medium, and difficult radar symptoms for the students to troubleshoot. Students worked with partners to accomplish tasks on the radar, and then, rotated to MACH III or to the paper-based method to perform additional tasks. The instructors recorded the task name, the date, the training method used (MACH III, paper, or radar), and the start time and end time.

Approximately 4 days were spent troubleshooting transmitter malfunctions and 3 days were spent troubleshooting the receiver. Instructors encouraged students to accomplish as many tasks on the troubleshooting list as possible. Also, instructors determined which tasks should be done and in what order, based on their teaching experience and the constraints of the rotation scheme. Students worked with "easy" tasks at first, but eventually progressed to the "medium" and "difficult" tasks of each circuit.

Standard practical examinations for the course required students to identify transmitter and receiver parts and describe their function. They also required students to perform check procedures, symptom recognition, troubleshooting, and signal tracing. Students could use all pertinent manuals and schematics, but were limited to 45 minutes for the total examination. To control for ceiling effects, additional practical examinations contained troubleshooting problems of greater difficulty and specific malfunctions that the students had not seen before. The problems did not coincide step-for-step with the manual. Students were required to think on their own and fill in logical steps where the manual left a gap.

The standard paper-and-pencil examination covered 20 multiple-choice transmitter questions randomly selected from a test question databank. The examination prompted students about where to find answers in the manuals and schematics; however, students had one hour to complete the test. To control for ceiling effects, additional paper-and-pencil examinations covered 20 transmitter and 20 receiver questions hand-selected from the databank

for greater difficulty. These latter examinations gave no prompts as to where answers could be found and lasted one hour each. All examination measures included percent correct and completion times.

Results

Examination scores. Twenty-nine students participated in the training and standard examinations. However, only 27 students were present for the additional examinations. Table 1 presents the mean scores (\bar{X}) of the practical and the paper-and-pencil examinations for the two groups. Recall that the additional examinations were designed to be more challenging than the standard examinations. Generally, students scored lower on the additional examinations than on the standard examinations, regardless of group.

Three out of seven *F* tests showed that the variances (s^2) of the MACH III group distributions were significantly smaller than those of the paper-based group. Significance differences in variance meant that the standard deviations (*s*) differed significantly as well. The differences made important implications for the effectiveness of MACH III. They indicated that students of the MACH III group performed more consistently than students of the paper-based method of instruction.

Table 1 Mean scores and standard deviations of examinations.

Examinations	Paper-Based Group			MACH III Group			Test for s^2 Differences
	n	$\bar{X}\%$	s	n	$\bar{X}\%$	s	F
TRANSMITTER							
Std. Practical	14	93	18.6	15	95	5.1	13.49*
Added Practical	13	71	32.9	14	75	24.9	1.75
Std. Paper & Pencil	14	85	8.8	15	88	8.2	1.15
Added Paper & Pencil	13	79	10.0	14	79	5.1	1.0
RECEIVER							
Std. Practical	14	93	18.1	15	99	2.1	71.42*
Added Practical	13	95	11.8	14	98	3.6	10.69*
Added Paper & Pencil	13	66	13.0	14	70	10.8	1.45

*Significance, $p < .05$.

The seven sets of examination scores were converted to *z* scores to meet the assumption of equal variance prior to performing analyses of variance. When comparing the groups, the MACH III group tended to have higher examination scores than the paper-based group, but univariate analyses of variance performed on the *z* scores showed that these differences were not statistically significant. A multivariate analysis of variance was also performed on the *z* scores of all seven examinations, but Pillai's Trace was not significant. Also, Mann-Whitney *U* tests on the ranked times showed no significant group differences based on completion times.

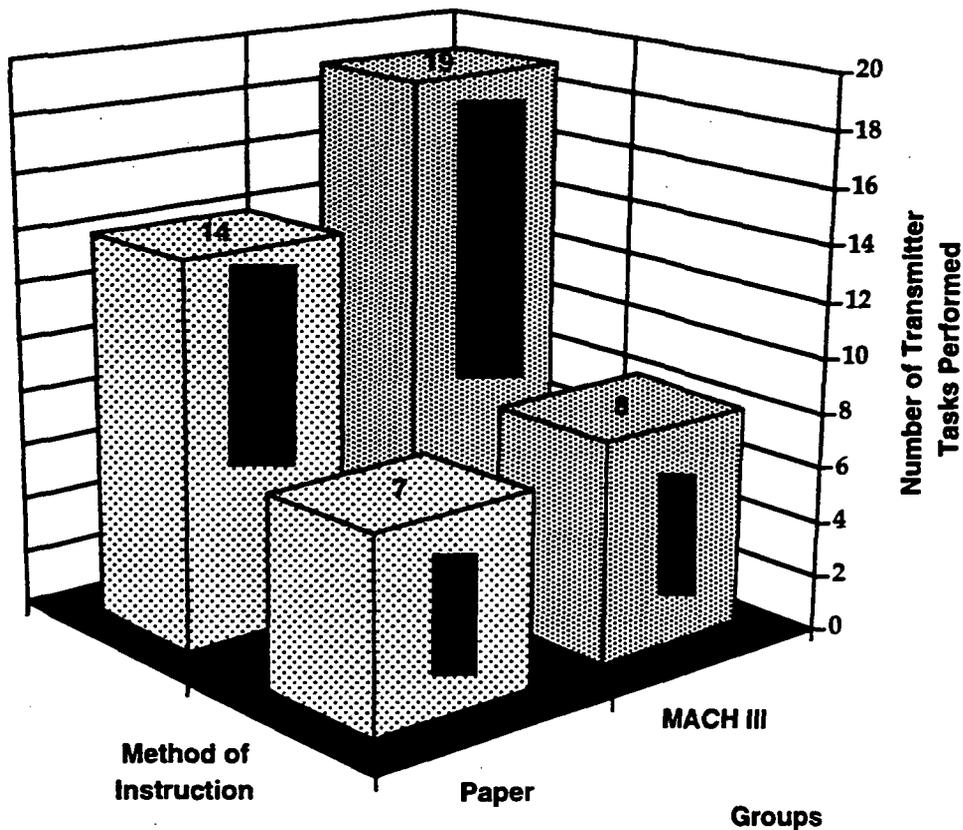


Figure 1. Mean number of transmitter tasks performed.

Number of tasks performed. The troubleshooting task list contained 25 transmitter tasks and 26 receiver tasks. In many cases, students completed the tasks for that day's subject area early, so additional tasks were assigned. Progress on these additional tasks was still recorded by when and how they were performed, but the tasks themselves were not identified. The following analysis is based on the number of tasks completed in drills by 14 paper-based students and 15 MACH III students.

The troubleshooting task list contained easy, medium, and difficult radar symptoms for the students to troubleshoot. During transmitter troubleshooting, the MACH III group averaged 13 easy and 13 medium/difficult tasks. By comparison, the paper-based group averaged 11 easy tasks and 9 medium/difficult tasks. Figure 1 shows that the MACH III group performed 1.4 times as many transmitter tasks using their supplemental method of instruction as the paper-based group did, $t(27) = 4.15, p < .05$.

During receiver troubleshooting, the MACH III group averaged 11 easy and 6 medium/difficult tasks. The paper-based group averaged only 8 easy tasks and 3 medium/difficult tasks. Figure 2 shows that the MACH III group performed 2.4 times as many receiver tasks as the paper-based group did using their supplemental instruction. The difference in number of tasks completed was significant, $t(27) = 8.05, p < .05$. These results indicate that the two methods of instruction were vastly different in terms of training efficiency. Not only did the MACH III group receive more troubleshooting practice, they received more challenging practice because of the efficiency of MACH III.

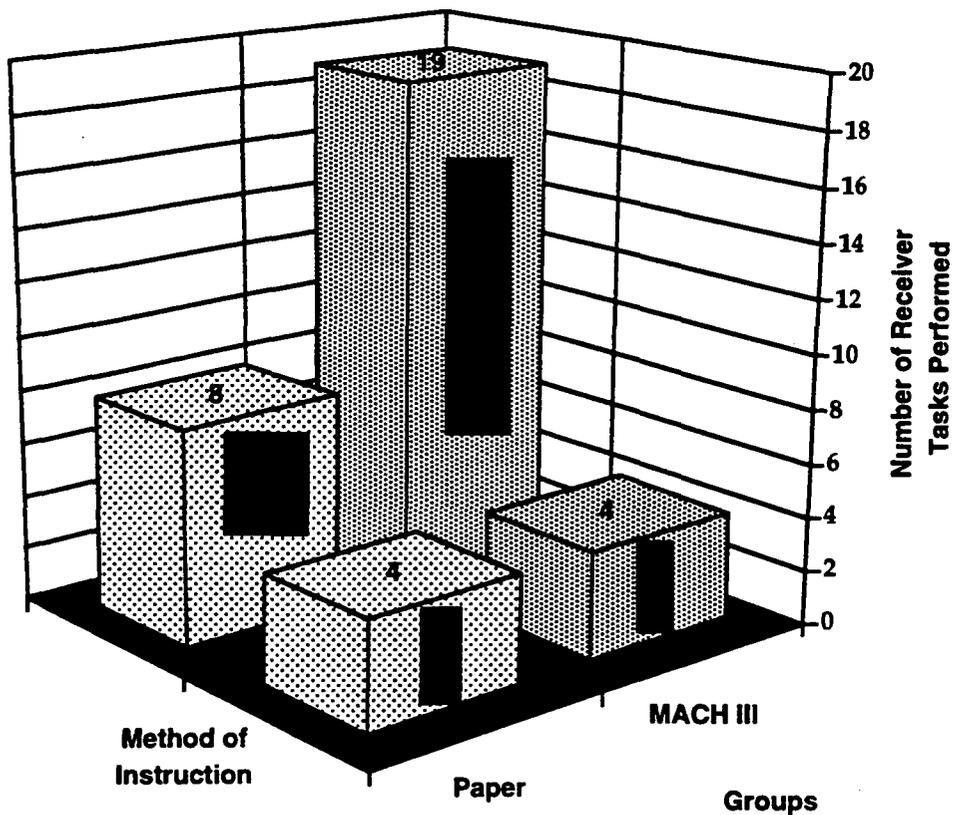


Figure 2. Mean number of receiver tasks performed.

Time spent between tasks. Because the two groups differed so much in amount of practice, the length of time between tasks was examined. Analysis indicated the median time between tasks was 8 minutes for the MACH III group and 30 minutes for the paper-based group. The time delay resulted from the paper-based group's dependence on outside sources for task information. The students who sat at the desk depended on their peers at the radar for information about meter readings, lamp lights, etc. Until students at the radar had progressed far enough in a task to determine certain information, students who troubleshooted on paper could not proceed. Once the students on the radar relayed the information, the students at the desk tended to complete the task very quickly. In fact, the students who troubleshooted on paper often sat idle for several minutes while the students on the radar finished the actual troubleshooting procedures.

By contrast, the students on MACH III worked independently from their peers on the radar. All the information needed to complete the tasks was contained in MACH III's simulation of the radar. Because of the dependency of the paper-based method on hands-on performance, the paper-based group progressed more slowly than the MACH III group did.

Time to perform tasks. Next, the time taken to complete troubleshooting tasks was examined. Recall that instructors determined which tasks should be done and in what order. Since instructors skipped around in the task list, the two groups did not always perform the same tasks. The following analysis reports the median time based on only those specific tasks which both groups performed. The analysis of common tasks included 14 transmitter tasks and 10 receiver tasks.

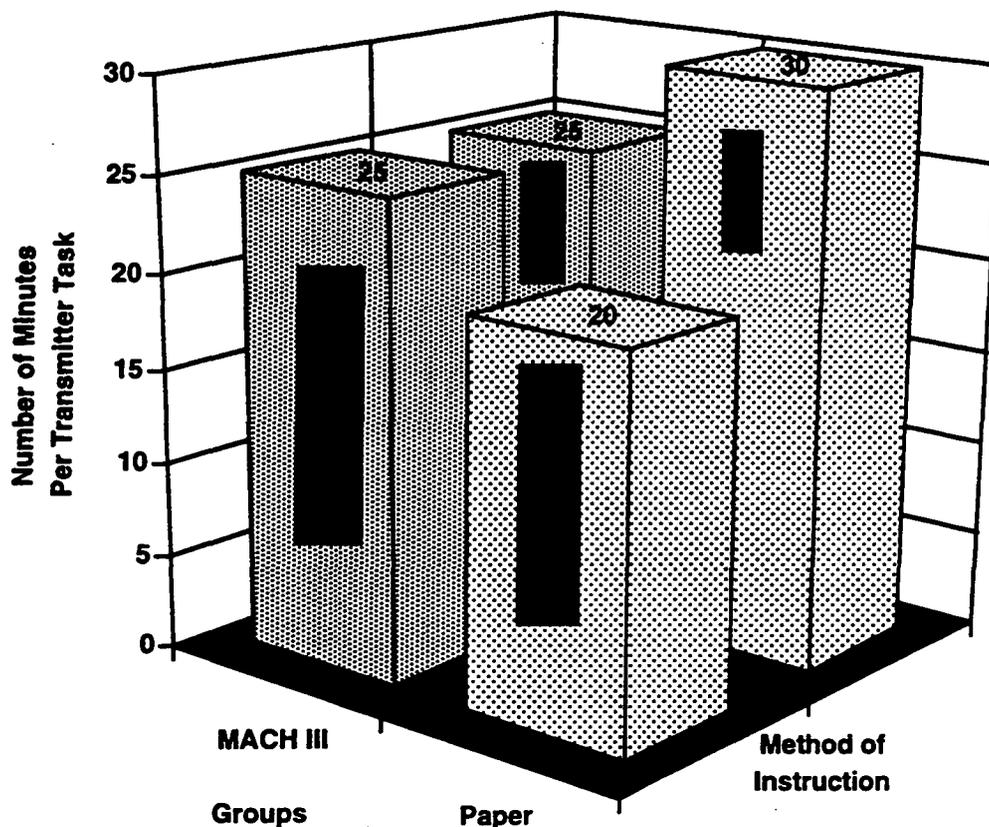


Figure 3. Median time to perform transmitter tasks.

Figure 3 shows that the paper-based group performed transmitter tasks on paper 5 minutes faster than the MACH III group did on MACH III. A Mann-Whitney \bar{U} Test of the ranked times was significant ($p < .05$). This savings occurred partly because paper-based students often took shortcuts in procedures. According to the instructors, once the students at the desk had received essential information from the radar, they were free to skip steps they had seen before.

In MACH III's software, however, the fault isolation procedures had to be followed, much like the actual procedures performed on the radar. Little, if any, opportunity existed for shortcuts or glossing over procedures. Each step in the procedures required an active response from the student. Furthermore, the troubleshooting times reflected the level of task difficulty. Medium and difficult tasks took longer to perform than easy tasks. Because MACH III students performed more medium and difficult tasks than paper-based students, their average times were longer. Together, the difficulty level and the lack of shortcuts slowed performance on MACH III.

Performance time on the radar's transmitter was a different story. The MACH III students performed transmitter tasks 5 minutes faster than the paper-based group. A Mann-Whitney \bar{U} Test of the ranked times was significant ($p < .05$). The MACH III students had performed so many tasks on the MACH III they were likely to perform similar tasks on the radar itself. The resulting practice effect may have helped MACH III students isolate faults more quickly than the paper-based students. Further, the students may have applied new, more efficient troubleshooting strategies they had learned through MACH III.

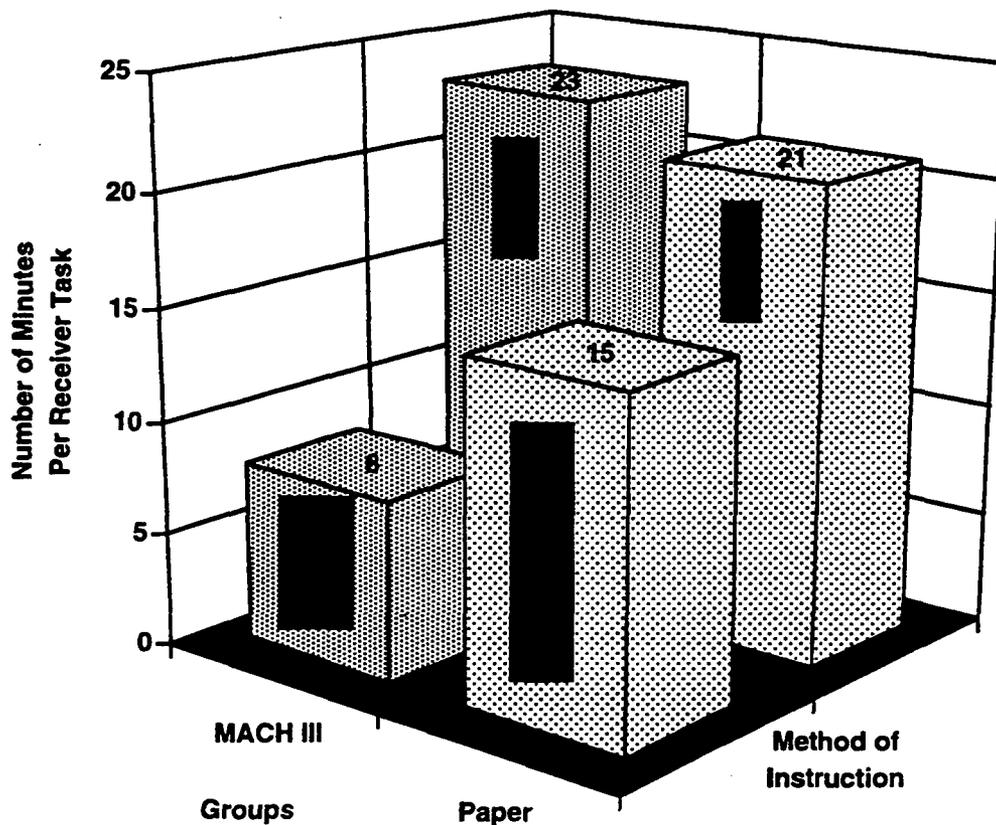


Figure 4. Median time to perform receiver tasks.

Figure 4 shows that the MACH III group performed receiver tasks on MACH III 7 minutes faster than the paper-based group did on paper. A Mann-Whitney U test of the ranked times was significant ($p < .05$). Instructors reported that once MACH III students had performed one lamp test on the receiver simulation, they could perform all remaining tasks without consulting the manuals. Also, the MACH III students probably had grown accustomed to using MACH III before receiver training began. As a result, they lost no time due to lack of familiarity with the training device. Figure 4 also shows that the paper-based group performed receiver tasks on the radar 2 minutes faster than the MACH III group did. This difference was not significant.

EVALUATION CONCLUSIONS

Overall, MACH III provided a more structured and time efficient method of instruction than the paper-based method. Given the same course length, more troubleshooting tasks were accomplished using MACH III. Although, more tasks did not produce higher examination scores, MACH III instruction resulted in greater consistency of performance overall and faster performance on the transmitter tasks.

Arguably, the same case can be made for the use of intelligent tutoring systems (ITS) in general. That is, an intelligent tutoring system can help to eliminate low scoring performances. Also, the greater task coverage afforded by an ITS can help students develop accurate and efficient mental models. When students attempt actual troubleshooting for the first time, their ITS practice can lead to a time savings on the real task.

Evaluation postscript. During the evaluation, instructors provided some insight into how MACH III might be revised. They noted that fault isolation checks required mechanics to walk to different sides of the radar to

monitor the status of lamps. However, the MACH III did not allow students to "view" the radar from all sides. As a result, an instructor suggested that future improvements to MACH III should more closely represent the dynamic and visual aspects of such procedures. Unknowingly, the instructor had voiced the demand for a virtual training environment.

VIRTUAL REVISIONS

REFERENCES

Acchione-Noel, S. C. (1991a) The Benefits of Training Radar Mechanics with an Intelligent Tutoring System. Rapallo (Genova), Italy: Proceedings of the Sixth International PEG Conference: Knowledge Based Environments for Teaching and Learning.

Acchione-Noel, S. C. (1991b) Intelligent tutoring: Effective and efficient training for radar mechanics. Fort Lee, VA: Proceedings of the Army Operations Research Symposium.

Acchione-Noel, S. C., Saia, F. E., Williams, L. A., & Sarli, G. G. (1990) Maintenance Aid Computer HAWK Intelligent Institutional Instructor Training Developments Study, TRAC-WSMR-TEA-90-051. White Sands Missile Range, NM: U.S. Army TRADOC Analysis Command-WSMR.

Pstotka, J., Massey, L. D., Mutter, S. A. (Eds.) (1988) Intelligent Tutoring Systems: Lessons Learned. Hillsdale, N. J.: Lawrence Erlbaum Associates Publishers, pp. 451-478.

Hollan, J. D., Hutchins, E. L., & Weitzman, L. (Summer 1984) STEAMER: An Interactive Inspectable Simulation-Based Training System. The AI Magazine, pp. 15-27.

BIOGRAPHY

Sylvia Acchione-Noel has conducted training effectiveness analyses for the US Army since 1981. She can be reached at US Army TRAC-WSMR, ATTN: ATRC-WGB, White Sands Missile Range, NM 88002-5502. DSN 258-5065. Commercial phone: (505) 678-5065. E-mail address: acchions@wsmr-emh91.army.mil.