Empirical evaluation of the transfusion medicine tutor

J.H. Obradovich, P.J. Smith, S. Guerlain, S. Rudmann, P. Strohm, J. Smith, J. Svirbely, and L. Sachs

Previous research during the development of Antibody IDentification Assistant (AIDA) revealed that many medical technology students and other laboratory personnel have serious difficulties in determining the specificity of blood group alloantibodies, especially weak or multiple antibodies. Based on these previous results, AIDA was modified to provide a teaching environment for medical technology students. We report the results of a rigorous, objective evaluation of the resultant system, the Transfusion Medicine Tutor (TMT). The results show that the students who were taught by an instructor using TMT to provide the instructional environment went from 0 percent correct on a pretest case to 87 percent correct on posttests (n = 15). This increase compares with an improvement rate of 20 percent by a control group (n = 15) who used a passive version of the system with the tutoring functions turned off. ***Immunohematology*** 1996;12:169–174.

Previous studies have shown that advances in computer technology can help train medical technology students and other laboratory personnel who have difficulties solving antibody identification cases, especially those cases involving weak or multiple antibodies.[1–3] Results on the performance of such personnel, when supported by an expert critiquing system, the Antibody IDentification Assistant (AIDA), showed that the use of this tool reduced errors by 30 percent to 62 percent on four posttest cases involving multiple antibodies.[4] The study presented here provides data regarding the effectiveness of the expert-system–based Transfusion Medicine Tutor (developed as an offshoot of AIDA) when used by medical technology students to develop skills in identifying alloantibodies in a patient's blood. The problem-solving context is described and important design features are explicitly identified. The results of a rigorous empirical evaluation also are discussed.

The Transfusion Medicine Tutor

Antibody identification is a laboratory task in which medical technologists perform a series of tests and analyze data to identify blood group antibodies in serum or plasma. This task has the classical characteristics of an abductive reasoning task, i.e., reasoning to the best explanation of the data. Some of the important general characteristics include:

- Multiple primitive solutions can be true at the same time (i.e., the patient could have several such antibodies).
- One or more antibodies can mask, or cover up, the presence of another antibody.
- The data are noisy (i.e., the quality of the data are questionable).
- The data are costly, so it is important to carefully select which tests should be run.
- There can be time stress to complete the task.
- The solution space is very large (a large number of alloantibodies have been identified).

To teach medical technology students about antibody identification, a problem-solving computer-based environment was developed in which students could solve actual patient's cases by requesting and interpreting test results. Figures 1 and 2 illustrate screen displays from this system for accessing and viewing data.

The Transfusion Medicine Tutor (TMT) has an expert system that monitors a student's performance for evidence of commission errors (erroneous inference), omission errors (lacking an expected inference), errors due to incomplete protocols (insufficient converging evidence), and errors due to lack of consistency with the available data.[5–7] Figures 3 and 4 illustrate the feedback provided in response to an error of commission and a final answer that is inconsistent with the data.

The TMT was explicitly designed to support teaching by a human. Its interface design improves the instructor's ability to diagnose a student's problems and misconceptions by making a student's thought processes more explicit to the teacher.[8,9] The teacher is able to use the screen to understand where in the task the student is encountering difficulty. The teacher is also able to pull up a summary of the student's work on each case, allowing for a more complete diagnosis of the student's thought process and progress.

Originally, TMT consisted only of complete cases for

Fig. 1. Test selection menu.

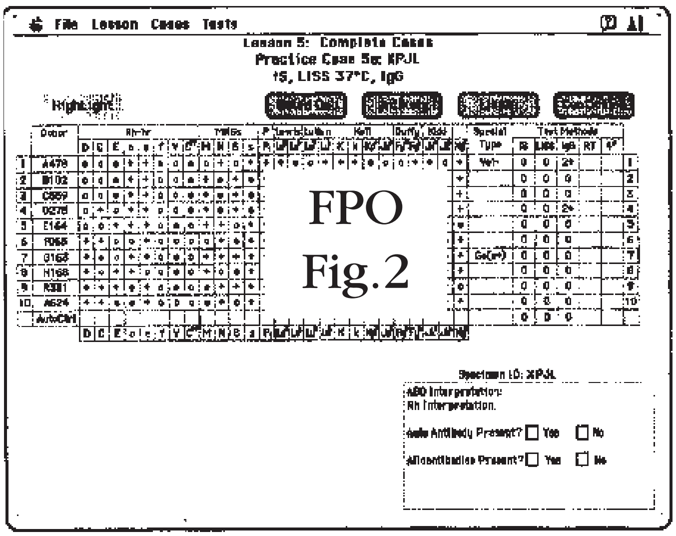

Fig. 2. Sample test panel with marker buttons.

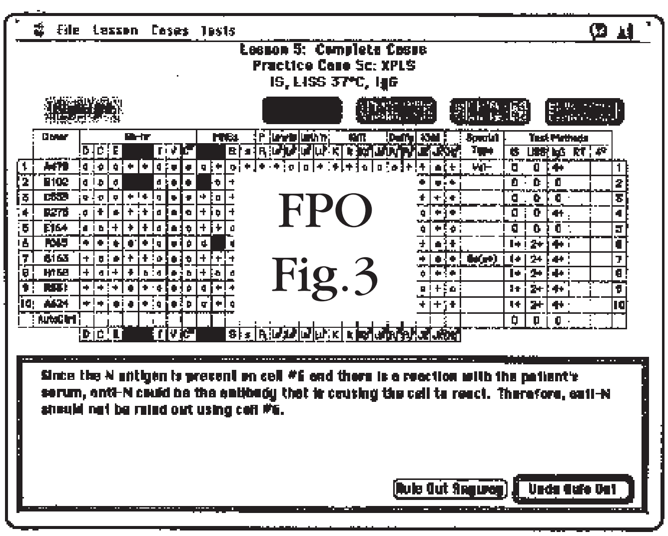

Fig. 3. Sample error message for an invalid rule-out (an error of commission).

the students to solve. An initial formative evaluation produced results that were disappointing. Many of the stu-

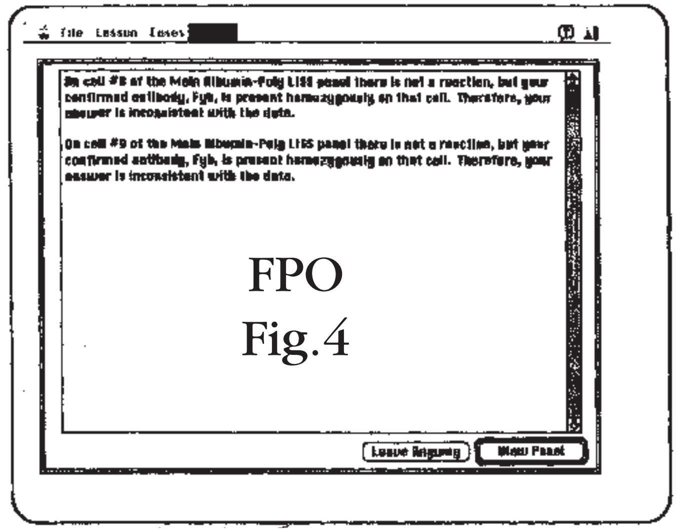

Fig. 4. Sample error message for a final answer that is inconsistent with the data.

dents seemed overwhelmed by the amount of material that TMT was attempting to teach (even though the students had already covered this material in their usual course work). Because of these results, five lessons were designed to teach the students a number of the critical subtasks before asking them to solve a complete case.[10,11]

Based on initial formative evaluations, we also concluded a computer system that is only reactive to a student's errors may not be the most efficient or effective method of teaching (at least with the extensive number of difficulties these students encountered). To remedy this situation, we designed a checklist that made explicit the high-level goals embedded in TMT's expert model and that summarized some of the key knowledge necessary to achieve these goals. The checklist serves several purposes:

- It acts as a reminder of certain types of knowledge necessary for antibody identification.
- It acts as a representation of the kinds of knowledge the expert system is expecting the user to apply in solving a case.

An Empirical Evaluation

This study explored the question:

What is the impact on learning in an environment where a human teacher uses the computer to provide a problem-based learning environment and to assist with teaching by providing context-sensitive feedback?

Procedure

Thirty students in the Medical Technology Program at a major university in the United States were tested using TMT. These students, college juniors, had completed the didactic and laboratory portions of their immunohematology course work, but had not yet begun their clinical rotation. Participation in the study was voluntary, but the students were paid for their participation. The study was conducted at another university where the staff had not been involved in the development of the system.

Experimental design

Half of the participants were randomly assigned to the control group and half to the treatment group. All participants were tested on the same cases in the same order with the exception of the pretest case and first posttest case. These two cases were randomized with respect to their order of use for each student. The first two matched cases were made to look like a single antibody when, in fact, two other antibodies were generating the pattern of reactions. All participants went through the same lessons in the same order, with the same practice cases and quizzes.

The first case (training case 1) was used to give both groups the same initial training on the use of the system's interface. This training was identical for both groups (the computer's intelligence was turned off for the treatment group).

The second case was used as a pretest case, allowing the experimenter to get a benchmark on a student's current performance in solving antibody identification and to identify the problem-solving strategies used prior to tutoring. This pretest case was performed by both groups with the system's critiquing functions turned off, and was matched in characteristics with the first posttest case.

Following the pretest case, the students completed five lessons; each of the first four lessons consisted of subtasks used to solve a complete case. The fifth lesson, composed of complete cases, consisted of solving a patient's complete case and included the use of all the subtasks covered in the first four lessons, along with more comprehensive strategies for gathering converging evidence to test a hypothesis.

Experimental treatment

The control group and the treatment group were introduced differently to the lessons.

Control group. Following the pretest case, the students in the control group were shown an example of the type of summary they would see at the end of each case. This summary pointed out any differences between their answers and the correct answer. It also described how an expert would have solved the subtasks or case. Thus, the control group had access to a passive tutor that only provided messages at the end of a subtask or case. After being shown this example summary, the control group began the lessons without further instructions. This group did not receive immediate feedback from the TMT during the course of solving the subtasks or cases in each lesson. Only by comparing their final answers with the system's answers, and by utilizing and processing the information in the summary at the end of a case, could they learn what errors they had made. Students would then use that information to avoid making the same errors in solving future cases.

Treatment group. Following the pretest case, the students in the treatment group were introduced to the intelligent tutoring provided by the system. During the first lesson, they were instructed to purposely make an error in order to experience the kind of intelligent feedback the computer would provide. The students were told to read the error message and to note that the message included the reason for their error. Like the control group, these students were also shown the information contained in the summary available at the end of each case.

In addition to the immediate, context-sensitive tutoring provided by the computer, treatment group students were given a paper checklist detailing the high-level goal structure that guided the expert system's error detection and tutoring. This checklist consisted of the step-by-step procedure the students should follow to successfully solve the cases contained in the tutorial. Students were told that different steps in the checklist corresponded to different lessons, and that the fifth lesson included using all the steps that were appropriate for solving each antibody identification case.

The students in the treatment group also had access to a human instructor for any questions or difficulties that they encountered while they worked through the lessons.

Treatment differences. As described above, the control group represented a very passive system but one that nevertheless provided students with access to a full description of expert performance on each case. The treatment group, on the other hand, differed in four ways from the control group (immediate, context sensitive tutoring by the computer, access to the checklist, and access to a human instructor). Thus, the treatment group represented an attempt to provide a "best-case" environment for teaching students, using the tutoring system to provide a learning environment and to provide active tutoring to assist the instructor's activities.

Posttest cases. Following the fifth lesson (complete cases), two posttest cases were given to all students. The first case matched the pretest case. The second case (in which one antibody masked a second that was also present) was the same for all subjects. For the two posttest cases, the computer's intelligence was turned off and end-of-case summaries were not provided. Students were, however, allowed to use the checklist for each posttest case.

Questionnaire. Following use of the system, a questionnaire was administered to each student to gather demographic data, to assess the students' subjective reactions to TMT and its various functions, and to elicit suggestions for improvement in the TMT design and use.

Results

The results showed that there were no significant differences in the misidentification rates on the pretest case for the control and treatment groups (see Table 1). However, the students in the treatment group showed a significant ($p < .001$) improvement in performance (a reduction from 100% to 13% misdiagnosis error rate) from the pretest case to the matched posttest case (evaluation of case 1). Students in the control group showed a 20 percent reduction in errors that was not a statistically significant improvement from the pretest to posttest case 1.

The between-subject analysis showed significant differences in performance ($p < 0.01$) on the posttest cases between the two groups (see Table 1). On posttest case 1, subjects in the treatment group had a misidentification rate of 13 percent, while subjects in the control group had a misidentification rate of 73 percent. On posttest case 2, students in the treatment group had a 7 percent misidentification rate, while students in the control group had a 73 percent misidentification rate. Thus, something about the treatment group (the use of intelligent tutoring, the checklist, and/or instructor assistance) produced a sizable and statistically significant improvement in performance.

Table 1. Misidentification rates for students in the treatment group (n = 15) versus the control group (n = 15)

Group	Pretest	Posttest case 1	McNemar's chi square	Posttest case 2
Treatment	15/15 wrong (100%)	2/15 wrong (13%)	X^2=11.0769 S @ $p < 0.001$	1/15 wrong (7%)
Control	14/15 wrong (93%)	11/15 wrong (73%)	X^2=1.3333 NS @ $p < 0.05$	11/15 wrong (73%)
Fisher's exact test	p = 0.50 NS @ $p < 0.05$	p = 0.0013 S @ $p < 0.01$		p = 0.0002 S@ $p < 0.001$

NS = Not significant
S = Significant

Classes of errors

In order to better understand the impact of the treatment condition on learning, we used the computer logs to identify error frequencies for five classes of errors (see Table 2). On the pretest, there were no significant differences between the control group and the treatment group. On the matched first posttest case, errors 2, 3a, 3b, and 4b each showed significant differences ($p < 0.05$) between the two groups, with the treatment group making fewer errors (see Table 2). Thus, tutoring appeared to be effective for errors that the computer could detect during the process of solving a case, as well as for errors that were detected after the student marked a final answer for a case.

Questionnaire results

Finally, the questionnaire results provided valuable, supportive data about students' perceptions regarding the system's usefulness and usability (see Table 3).

Of the students in both groups, 100 percent agreed or strongly agreed with the statement "I would recommend this program to other students," and also to the statement "The program would be useful as a study tool." Also, of the students in both groups, 97 percent agreed or strongly agreed with statements 1, 4, and 6 (see Table 3). It is important to note how insensitive such questionnaire results are to actual performance differences between the two groups.

The instructor's responses to the open-ended questions are shown in Table 4. Some of the instructor's responses are summarized below:

"What did you like best about the tutorial?"

"Panels—highlighting cells; rule-out; computer procedures; the flow of the program."

"How would you improve this tutorial?"

"Exception rules for marking low-frequency antibodies are too difficult for this student level."

"I look forward to getting your IBM version of this tutorial."

Conclusion

Although the relative contributions of the computer versus teacher versus checklist cannot be determined from this study, the results provide strong evidence that an effective learning environment was developed. The results, when combined with previous studies using traditional teaching methods, suggest that use of such

Table 2. Classes of errors made by the treatment group and control group participants on the pretest and posttest case 1

Error	Subjects committing error at least once on pretest case 1: Treatment group (n = 15)	Control group (n = 15)	Fisher's exact test	Subjects committing error at least once on posttest case 1: Treatment group (n = 15)	Control group (n = 15)	Fisher's exact test
1. Ruling out correct answer due to ruling out incorrectly	7	5	p = 0.3553 NS	2	4	p = 0.3257 NS
2. Failure to rule out when appropriate	13	13	p = 0.7011 NS	5	11	p = 0.0328 S
3. Failure to collect converging evidence						
a. Failure to do antigen typing	9	8	p = 0.5000 NS	1	8	p = 0.0070 S
b. Failure to satisfy the 3+/3– rule	7	6	p = 0.5000 NS	1	6	p = 0.0401 S
4. Failure to check for consistency of data with answer						
a. Failure to ensure there are no unexplained negative reactions	1	3	p = 0.2988 NS	1	2	p = 0.5000 NS
b. Failure to ensure there are no unexplained positive reactions	14	11	p = 0.1648 NS	2	8	p = 0.0251 S

S = Significant ($p < 0.05$)
NS = Not significant ($p > 0.05$)

Table 3. Student questionnaire results

Statements about computer tutorial	Version	SD	D	N	A	SA
1. The program was easy to use.	Control	–	–	–	4	11
	Treatment	–	–	1	5	9
2. I learned a great deal from the program.	Control	–	–	2	4	9
	Treatment	–	–	1	8	6
3. The program would be useful as a study tool.	Control	–	–	–	2	13
	Treatment	–	–	–	5	10
4. The lessons contributed to my understanding of the topic.	Control	–	–	–	4	11
	Treatment	–	–	1	5	9
5. The program was too long.	Control	3	4	5	3	–
	Treatment	2	6	7	–	–
6. The instructions were clearly written.	Control	–	–	1	6	8
	Treatment	–	–	–	8	7
7. I would recommend this program to other students.	Control	–	–	–	3	12
	Treatment	–	–	–	3	12
Statements about checklist (control group only)						
8. The checklist was easy to use.	Treatment	–	1	–	8	6
9. The checklist was well organized.	Treatment	–	–	–	7	8
10. The checklist was useful.	Treatment	–	1	–	8	6

SA = Strongly Agree
D = Disagree
N = Neutral
A = Agree
SD = Strongly Disagree

Table 4. Instructor questionnaire results

Statements about computer tutorial	SD	D	N	A	SA
1. This tutorial provided a useful teaching strategy.	–	–	–	–	✓
2. The instructions were clearly written.	–	–	–	–	–
3. The software was easy to use.	–	-	–	–	✓
4. This tutorial helped me learn about how students solve problems.	–	–	✓	–	–
5. This experience will cause me to teach differently in the future.	–	–	–	✓	–
6. This tutorial could be useful for student self-study.	–	–	–	–	✓
7. I would like to incorporate this tutorial into my classroom.	–	–	–	–	✓

SA = Strongly Agree
D = Disagree
N = Neutral
A = Agree
SD = Strongly Disagree

technology could significantly enhance the quality of education for medical technology students.

Acknowledgments

This work has been supported under a grant (HL51611-03) from the National Heart, Lung, and Blood Institute, Bethesda, MD. The authors thank Veronica Lewis of the University of Illinois (Chicago) for her assistance in conducting this study.

References

1. Miller T, Smith PJ, Gross S, et al. The use of computers in teaching clinical laboratory science. Immunohematology 1993;9:22–7.
2. Smith PJ, Galdes D, Fraser J, et al. Coping with the complexities of multiple-solution problems: a case study. Intern J Man-Machine Studies;35:429–53.
3. Smith PJ, Miller T, Fraser J, et al. An empirical evaluation of the performance of antibody identification tasks. Transfusion 1991;31:313–7.
4. Guerlain S, Smith PJ, Obradovich JH, Smith J, Rudmann S, Strohm P. Proceedings of the 1995 IEEE International Conference on Systems, Man, and Cybernetics. The Antibody IDentification Assistant (AIDA): an example of a cooperative computer support system, 1995:1909-14.
5. Burton, RR. Diagnosing bugs in a simple procedural skill. In: Sleeman D, Brown JS, eds. Intelligent Tutoring Systems, London: Academic Press, 1982:157–84.
6. Fox B. The human tutorial dialogue project: issues in the design of instructional systems. Hillsdale, NJ: Lawrence Earlbaum, 1993.
7. Lojoie S, Lesgold A. Apprenticeship training in the workplace: computer-coached practice environment as a new form of apprenticeship. Machine-Mediated Learning 1989;3:7–28.
8. Bailey GD, ed. Computer-based integrated learning systems. Englewood Cliffs, NJ: Educational Technology Publications, 1993.
9. Interactive instruction and feedback. Demsey JV, Sales GC, eds. Englewood Cliffs, NJ: Educational Technology Publications, 1993.
10. Gagne RM. The conditions of learning and theory of instruction. 4th ed. New York: Holt, Rinehart, and Winston, 1985.
11. Gordon SE. Systematic training program design: maximizing effectiveness and minimizing liability. Englewood Cliffs, NJ: Prentice Hall, 1994.

Jodi Heintz Obradovich, Cognitive Systems Engineering Laboratory, The Ohio State University, 210 Baker Systems, 1971 Neil Avenue, Columbus, OH 43210-1271; Philip J. Smith, Stephanie Guerlain, Sally Rudmann, Patricia Strohm, Jack Smith, John Svirbely, and Larry Sacks, Cognitive Systems Engineering Laboratory, The Ohio State University, Columbus, OH.