The Transfusion Medicine Tutor: The Use of Expert-Systems Technology to Teach Students and Provide Support to Practitioners in Antibody Identification

Jodi Heintz Obradovich, Stephanie Guerlain, Philip J. Smith, Jack W. Smith, Jr., Sally Rudmann, Larry Sachs, John Svirbley, and Melanie Kennedy, and Patricia L. Strohm

Cognitive Systems Engineering Laboratory
The Ohio State University
210 Baker Systems, 1971 Neil Ave.
Columbus, OH 43210
jobradov@magnus.acs.ohio-state.edu (614) 292-1296
psmith@magnus.acs.ohio-state.edu (614) 292-4120

Abstract: The studies presented here provide data regarding the effectiveness of the expert-system-based Transfusion Medicine Tutor (TMT) when used by medical technology students and medical technologists in performing an important problem-solving task, the identification alloantibodies in a patient's blood for the purpose of finding compatible blood for transfusion. TMT monitors for slips and mistakes during the evaluation of patient cases, and provides suggestions to the student or technologist when a potential concern is detected. Use of that system was found to reduce errors in final answers on various multiple antibody cases by 30%-62% by practicing medical technologists. Similarly, use of a modified version of the system for teaching was found to enhance learning by 87-93%.

INTRODUCTION

We have conducted a series of empirical studies aimed at understanding how to engineer computer systems to successfully aid students and practicing medical technologists in learning more effective strategies to solve complex problems [Smith, Smith, Svirbely, et al. 1991; Smith, Miller, Fraser, et al. 1991]. The studies presented here provide data regarding the effectiveness of the expert-system-based Transfusion Medicine Tutor (TMT) when used by medical technology students and medical technologists in performing an important problem-solving task, the identification alloantibodies in a patient's blood for the purpose of finding compatible blood for transfusion.

ANTIBODY IDENTIFICATION

One of the most challenging problem-solving tasks confronted by medical technologists in transfusion services laboratories is determining what antibodies are in a patient's blood. Antibody identification is a laboratory workup task where medical technologists must run a series of tests and analyze a large amount of data to detect antibodies in a patient's blood (as part of the process for determining compatible blood for a transfusion).

This task exhibits the classic characteristics of an abductive reasoning task, i.e., reasoning to the best explanation of the data. Some of the characteristics found in the antibody identification task are: noisy data (the quality of the data is questionable), one antibody masking (or covering up) another antibody, costly data, a large solution space (more than 400 alloantibodies have been identified), and having multiple primitive solutions true at the same time (there may be multiple antibodies present in a patient's blood).

BACKGROUND STUDIES

This project began by conducting a series of studies aimed at understanding how to design expert-systems technology that can be applied to education. These studies led to the development of an expert model of problem solving, the identification of common errors and misconceptions in solving such problems, and development of a model of expert tutoring in the domain of antibody identification. The results of these background studies have been the basis for the development of an expert system that serves two areas of application:

1. Providing a computerized support tool that helps practitioners solve cases, while providing them with embedded training as they solve these cases;
2. Aiding medical technology students in developing their problem-solving skills for the task of antibody identification.

Empirical evaluations were conducted for both applications and are presented in this paper.

THE PROBLEM-SOLVING TOOL

To teach medical technology students about antibody identification and to provide embedded training and support to medical technologists, a problem-solving environment was developed in which students and practitioners could solve actual patient cases, requesting test results and interpreting these results in order to arrive at an answer.

This system, the Transfusion Medicine Tutor (TMT), has embedded in its architecture an expert system that monitors the student's performance [cf., Burton 1982; Fox 1993; Lajoie & Lesgold 1989] for evidence of errors of commission (erroneous inference), errors of omission (missing an expected inference), and errors due to incomplete protocols (insufficient converging evidence).

This system was designed to support teaching by a human teacher. In particular, its interface design improves the instructor's or supervisor's ability to diagnose students' or practitioners' problems and misconceptions by making more explicit to the teacher/supervisor the student/practitioner's thought processes [cf., Bailey 1993; Dempsey & Sales 1993]. The teacher/supervisor is given warnings as students make errors and is able to access displays indicating where in the task the student/practitioner is encountering difficulty in order to provide immediate assistance to the student.

Originally, TMT consisted only of complete patient cases for the student/practitioner to solve. An initial formative evaluation using medical technology students produced results that were disappointing. Many of the students seemed overwhelmed by the amount of material that TMT was attempting to teach (even though the students had already covered this material in their normal coursework). Because of these results, five initial lessons were designed that are intended to teach the students a number of the critical subtasks before asking them to solve a complete case [cf., Gagne 1985; Gordon 1993].

In addition, based on initial formative evaluations, we also concluded that a computer system that is only reactive to a student/practitioner's errors may not be the most efficient or effective method of teaching or support (at least with the extensive number of difficulties these students were encountering). 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:

- As an external memory aid, acting as a reminder of certain types of knowledge necessary for antibody identification.
- As a representation of the kinds of knowledge the expert system is expecting the user to apply in solving a case.

TMT AS SUPPORT TOOL FOR PRACTITIONERS

One purpose of this support tool was to study the effects that critiquing systems (where users perform a task and the computer uses its knowledge to critique the person's performance) can have on the performances of technologists when identifying the alloantibodies present in patients' blood. This system monitors for slips and mistakes made during the evaluation of patient cases and provides intelligent, context-sensitive feedback to the technologist when a potential concern is detected.

The Practitioners

The background studies involving practitioners engaged in the antibody identification task have provided evidence that practitioners make a significant number of errors [Smith, Smith, Svirbley, et al. 1991; Strohm, Smith, et al. 1991; Guerlain 1993] including slips, failure to form appropriate hypotheses, failure to collect independent converging evidence for the answer, failure to use all information available from a test result, and biased assimilation. Many of these errors are due to inadequate training and practice with the antibody identification task since the practitioner may only perform this task occasionally.

In this study, 32 practitioners from six hospitals were tested. All of these technologists were identified by their supervisors as "actually performing the task of antibody identification as part of their job, but who would benefit from additional experience and training." Their years of experience ranged from one to 35 years (with a mean of 10 years).

Experimental Design

The practitioners were randomly assigned to the Treatment or Control Group, the former using the checklist and a version of the system with all critiquing functions operative and the latter using the system without any feedback.

Performance was studied using a combination of a within- and between-subjects design. All practitioners were first trained on the use of the interface with the control version of the system. Following this training, they solved a Pre-Test Case without any aiding. This Pre-Test Case was matched with the first Post-Test case, allowing for the within-subjects comparison. Three additional Post-Test Cases were administered and were used in the between-subjects analysis. All of the cases had characteristics that made them fairly difficult to solve (multiple antibodies, one antibody masking another, noisy data, etc.) and were actual patient cases or realistic cases designed by an expert in antibody identification.

Results

Several important classes of errors are listed in Table 1. As indicated in this table, multiple process errors were made by the practitioners on the pre-test case. This finding is consistent with the evidence from previous studies that revealed that practicing medical technologists make a significant number of errors when solving antibody identification cases.

Error	# of practitioners
1. Ruling out hypotheses incorrectly	20
2. Failing to rule out when appropriate	14
3. Failure to collect converging evidence	26
4. Answer implausible given data	11
5. Answer implausible given prior probabilities	11

Table 1. Number of Practitioners (out of 28) who made process errors on the Pre-Test Case.

The analysis of misdiagnosis rates on the test cases (Table 2) showed that there was no significant difference in performance on the Pre-Test Case for the Control and Treatment Groups. A Log-Linear analysis was run to take into account the difference in performance between the two groups on the Pre-Test Case. This analysis gave an overall significance level of $p \leq 0.000003$, favoring performance for the Treatment Group.

Test Cases	Pre-Test Case	Post-Test 1	Post-Test 2	Post-Test 3	Post-Test 4
Control Group	6 out of 14 (43% wrong)	5 out of 15 (33.3% wrong)	8 out of 16 (50 % wrong)	6 out of 16 (37.5% wrong)	10 out of 16 (62.5% wrong)
Treatment Group	4 out of 15 (26.7% wrong)	0 out of 16 (0.0% wrong)	3 out of 16 (18.8% wrong)	0 out of 16 (0.0% wrong)	0 out of 16 (0.0% wrong)
Significance (between-subjects analysis)	NS	$p < 0.05$	$p = 0.067$	$p < 0.01$	$p < 0.001$

Table 2. Misdiagnosis Rates.

TMT AS TUTORING SYSTEM

Another design application of the Transfusion Medicine Tutor (TMT) is as a testbed for studying design concepts for the development of a general model for teaching abduction. The following study describes the results obtained when TMT was used as an expert-systems-based tutoring system for teaching medical technology students the abduction task of red cell antibody identification.

The Students

Thirty students in the Medical Technology Program at a major U. S. university were tested on TMT. These students were college juniors and had completed the didactic portion of their immunohematology coursework and the student lab. Participation in the study was voluntary, but the students were paid for their participation.

Experimental Design

Participants were randomly assigned to one of two groups: Half were assigned to be in the Control Group and half to the Treatment Group. The Treatment Group used the checklist, a version of the system with all critiquing functions operative, and the assistance of their instructor as needed. The Control Group used a version of the system that offered no immediate feedback from TMT, but provided a summary at the end of the case containing the student's answer for the case, the correct answer for the case, and how an expert would have solved the subtasks or cases.

All of the participants were trained on the interface using the same case (with all intelligent tutoring functions turned off) and were tested on the same cases in the same order with the exception of the Pre-Test Case and first Post-Test Case, which were randomized with respect to their order of use for each student (used in the within-subjects analysis). Like the training case, the version of TMT used for the Pre-Test Case provided no feedback .

Following the Pre-Test Case, the students completed five lessons; each of the first four lessons consisted of subtasks involved in solving a complete case. The fifth lesson, Complete Cases, consisted of solving complete patient cases, and included the use of all the subtasks covered in the first four lessons, along with more global strategies for gathering converging evidence to test a hypothesis.

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 three ways from Control Group:

- They received context-sensitive, immediate tutoring by the computer,
- The had access to the checklist, and
- They had access to a teacher.

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 in order to assist the instructor's activities.

Post-Test Cases. Following Lesson 5 (Complete Cases), all students completed two Post-Test Cases were given to all students. The first was one of the two randomly ordered, matched cases. The second case (in which one antibody masked a second that was also present) was the same for all students. For the two Post-Test Cases, the intelligence was turned off, no end-of-case summaries were provided and no instructor assistance was provided. The Treatment Group participants were, however, allowed to use the checklist for each post-test case.

Results

Misdiagnosis Rates. The results showed that there was no significant difference in the misdiagnosis rates on the Pre-Test Case for the Control and Treatment Groups (Table 3). 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 Pre-Test Case to the matched Post-Test Case. Students in the control group showed a reduction in error rate from 93% to 73% that was not statistically significant from the Pre-Test to Post-Test Case 1.

The between-subject analysis showed a significant difference in performance on the post-test cases between the two groups (Table 3). On Post-Test Case 1, students in the Treatment Group had a misdiagnosis rate of 13% while students in the Control Group had a misdiagnosis rate of 73%. On Post-Test Case 2, students in the Treatment Group had a 7% misdiagnosis rate, while students in the Control Group had a 73% misdiagnosis rate. Each of these differences is significant ($p < 0.01$).

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.

	Pre-test	Post-test Case 1	McNemar's Chi Square	Post-test Case 2
Treatment	15/15 wrong (100%)	2/15 wrong (13%)	$\chi^2 = 11.0769$ S @ $p < 0.001$	1/15 wrong (7%)
Control	14/15 wrong (93%)	11/15 wrong (73%)	$\chi^2 = 1.3333$ 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 @ $p < 0.001$

Table 3. Misdiagnosis rates for students in the Treatment Group (n=15) vs. the Control Group (n=15).

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 4). On the Pre-Test, there were no significant differences between the Control and Treatment Groups. On the matched first Post-test case, Errors 2, 3a, 3b, and 4b each showed significant differences ($p < 0.05$) between the Treatment and Control Groups, with the Treatment Group making fewer errors (see Table 4). 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.

	Subjects committing error at least once on Pre-test Case		Fisher's Exact Test	Subjects committing error at least once on Post-test Case 1		Fisher's Exact Test
Error	Treatment (n = 15)	Control (n = 15)		Treatment (n=15)	Control (n=15)	
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

Table 4. Classes of errors made by the Treatment and Control Group participants on the Pre-Test and Post-Test Case 1.

CONCLUSION

Although the application of expert systems technology to education has a very interesting intellectual history, its practical impact has been disappointing to date. Clearly, the task of antibody identification is a task that medical technologists find difficult, since they are getting moderately difficult, yet realistic, patient cases consistently wrong when unassisted. The studies presented in this paper illustrate the impact that a well-designed learning environment can have on the performance of medical technology students and practitioners in the task of antibody identification.

A systems approach was taken in the design of TMT, leading us to design a computer system that revolved around the application of a complete protocol, using a number of complementary problem-solving strategies to independently converge on an answer. In our study with practitioners, the critiquing model of interaction allowed the human practitioner to stay involved in the task, apply their own expertise, learn from the computer, and judge the computer's feedback in a context-sensitive manner. As a tutoring system, TMT provided tutoring as well as supporting teacher intervention. There was evidence that TMT aided both students and practitioners by catching slips and mistakes and helping users to recover from these errors, employing different types of error checking mechanisms (checking for errors of commission, checking for errors of omission, checking for an incomplete protocol, and for practitioners, checking that the answer was consistent with prior probability information).

The use of a checklist with both students and practitioners was beneficial in quickly training them on the high-level goal structure implicit in the computer's knowledge base, and served as a reminder to participants of the steps necessary to successfully solve a case. This checklist is meant to be used along with TMT (at least until the student or technologist internalizes the steps), so that the user has an appropriate mental model regarding the computer's expectations. The overall protocol involves collecting converging evidence to reduce the chances of errors in the final answer due to the use of fallible heuristics or human error.

The success of the system's interaction with the user relied on its unobtrusive interface that allowed both students and practitioners to solve antibody identification cases as they normally would using paper and pencil, while providing the computer with a rich set of data regarding the characteristics of the case and the user's problem-solving steps without requiring the student or practitioner to enter information that was outside the normal task requirements.

Although the relative contributions of the computer vs. teacher vs. checklist cannot be determined from these studies, 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 technology could significantly enhance the quality of education for medical technology students and provide valuable practice and support for practicing medical technologists. Additional studies are underway to identify the contribution of the various components of this environment to overall learning.

Acknowledgements

This research was supported by a grant from the National Heart, Lung, and Blood Institute.

References

[Bailey 1993] Bailey, G.D. (Ed.) (1993). Computer-Based Integrated Learning Systems. Englewood Cliffs, NJ: Educational Technology Publications.

[Burton 1982] Burton, R.R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman and J.S. Brown (Eds.), Intelligent Tutoring Systems, London: Academic Press, 157-184.

[Dempsey & Sales 1993] Dempsey, J.V. and Sales, G.C. (Eds.) (1993). Interactive Instruction and Feedback. Englewood Cliffs, NJ: Educational Technology Publications.

[Fox 1993] Fox, B. (1993). The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems. Hillsdale, NJ: Lawrence Earlbaum.

[Gagne 1985] Gagne, R.M. (1985). The Conditions of Learning and Theory of Instruction (4th Edition). New York: Holt, Rinehart and Winston.

[Gordon 1994] Gordon, S.E. (1994). Systematic Training Program Design: Maximizing Effectiveness and Minimizing Liability. Englewood Cliffs, NJ: Prentice Hall.

[Guerlain 1993] Guerlain, S. (1993). *Factors influencing the cooperative problem-solving of people and computers.*. Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, Seattle, WA. 387-391.

[Lajoie & Lesgold 1989] Lajoie, S.P. and Lesgold, A. (1989). Apprenticeship training in the workplace: Computer-coached practice environment as a new form of apprenticeship. Machine-Mediated Learning, 3, 7-28.

[Miller, Smith, et al. 1993] Miller, T., Smith, P.J., Gross, S., Guerlain, S., Rudmann, S. Strohm, P., Smith, J., and Svirbely, J. (1993) The use of computers in teaching clinical laboratory science. Immunohematology, 9, 22-27.

[Smith, Smith, Svirbley, et al. 1991] Smith, P.J., Smith, J., Svirbely, J., Galdes, D., Fraser, J., Miller, T., Blazina, J., Kennedy, M., Rudmann, S., and Thomas, D. (1991). Coping with the complexities of multiple-solution problems: A case study. International Journal of Man-Machine Studies, 35, 429-453.

[Smith, Miller, Fraser, et al. 1991] Smith, P.J., Miller, T., Fraser, J., Smith, J., Svirbely, J., Rudmann, S., Strohm, P., and Kennedy, M. (1991). An empirical evaluation of the performance of antibody identification tasks. Transfusion, 31, 313-317.

[Strohm, Smith, et al. 1991] Strohm, P., Smith, P. J., Fraser, J., Smith, J. W., Rudmann, S., Miller, T., & Kennedy, M. (1991). *Errors in antibody identification.* Immunohematology, 7, 20-24.