Quantitative Impact of a Cognitive Modeling Intelligent Tutoring System on Student Performance in Balancing Chemical Equations

Mary Beth Walsh,[†] Connie M. Moss,[‡] Benny G. Johnson,[§,*] Dale A. Holder,[§] and Jeffry D. Madura[†]

[†]*Department of Chemistry and Biochemistry and* [‡]*Department of Foundations and Leadership, Duquesne University, Pittsburgh, PA 15282;* [§]*Quantum Simulations, Inc., 5275 Sardis Road, Murrysville, PA 15668, johnson@quantumsimulations.com*

Received June 24, 2002. Accepted October 5, 2002.

Abstract: The need for improved interactive tutoring capabilities in educational software for chemistry problem solving is an important one clearly articulated by teachers and students. To deliver the next generation of individualized interactive capabilities users demand, it is necessary to go beyond the conventional computer-assisted instruction methodology. The focus of this paper is the assessment with first-semester general chemistry students of a recently developed artificial intelligence (AI) tutor for balancing chemical equations. This is the first such assessment of an AI-based learning tool in chemistry. Students in CHEM 121 in the Fall 2001 semester at Duquesne University (N = 273) participated in the study. Students were divided into a test group that used the AI tutor as part of their study activities and a control group that did not use the tutor. It was found that the tutor improved the performance of the test group students to a statistically significant degree, helping the weakest students the most. This study establishes the feasibility of an AI-based approach to creating advanced new tutoring software for chemistry problem solving. Access to a Web-based demonstration of the equation-balancing tutor may be obtained by emailing the corresponding author.

Introduction

The need for improved interactive tutoring capabilities in educational software for chemistry problem solving is an important one clearly articulated by teachers and students. Most current tutorial programs are termed “computer-assisted instruction” (CAI), an approach that has been used in chemistry for a long time [1, 2]. To deliver the next generation of individualized interactive capabilities users demand, however, it is necessary to go beyond this methodology.

We are currently engaged in a large-scale project for developing interactive artificial intelligence tutoring software for high school and college chemistry. The goal of this work is to incorporate new concepts from the field of artificial intelligence (AI) as a route to meaningful individualized tutoring, which CAI cannot deliver because of its intrinsically rigid design. The basic shift is to replace specific foreknowledge of problems and answers with a direct representation of chemical and pedagogical principles, and then simulate reasoning using these principles for the purpose of tutoring.

As part of this effort, we have recently developed a tutoring program for balancing chemical equations [3]. Balancing equations is a topic of considerable pedagogical interest, as evidenced by the large existing literature, which has recently been thoroughly reviewed [4]. The focus of this paper is the assessment of this tutor with first-semester general chemistry students. This is the first such assessment of an AI-based learning tool in chemistry.

The program studied here contains two important advances over conventional software. First, the system creates a worked-out solution with detailed explanations for *any* equation entered by the student or teacher. Unlike a conventional tutorial, this is done dynamically, without the equation being stored ahead of time. Second, the program interactively answers a variety of detailed questions for the student at each step in the solution. The particular pedagogical approach of this program is specifically oriented to help beginning and lower-performing students, who often cannot make any start on a problem or do not feel comfortable attempting to do so. For students who are ready to try problems for themselves, we have developed tutors that allow them to enter their own work and receive feedback and guidance.

The ability of the tutor to answer questions makes it possible for students to conduct exploratory inquiry even before they can attempt the problems. At each step, questions are displayed in a menu (depicted in Figure 1) from which students can choose as needed, selecting as many or as few questions as they desire. Many different paths of inquiry are possible for the same problem, with the student directing the inquiry [5]. The questions and answers are highly targeted and context-specific, changing at each step, and model good scientific thinking about the problem domain and underlying concepts. This is very important in fostering the development of the student’s own self-explanation [6] and question-asking [7] abilities. Support is provided on several different levels, from initial “hand-holding” to advanced conceptual questions. We do not take for granted, for example, that the student can answer elementary questions like “Is hydrogen balanced?” or even necessarily realize the relevance or importance of such

* Address correspondence to this author.
† Department of Chemistry and Biochemistry, Duquesne University
‡ Department of Foundations and Leadership, Duquesne University
§ Quantum Simulations.

 S1430-4171(02)06620-7, Published on Web 10/24/2002, 10.1007/s00897020620a, 760379bj.pdf

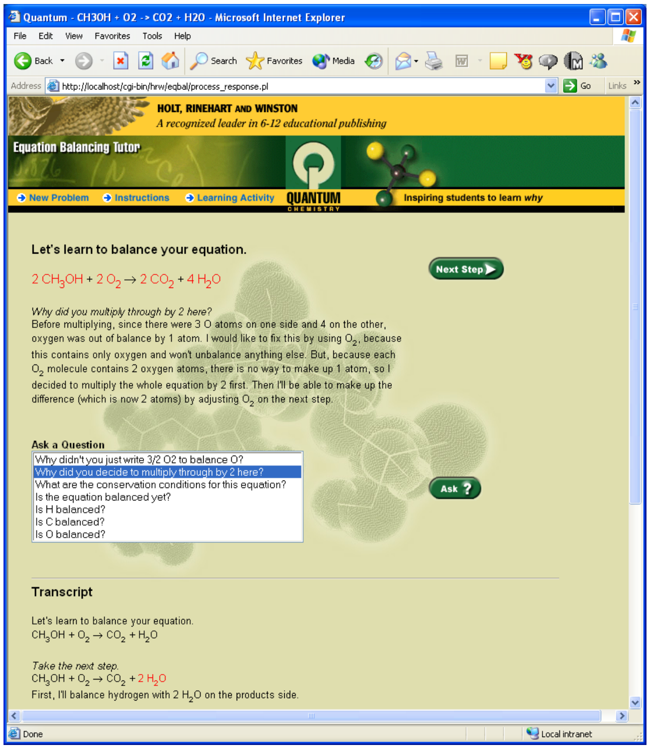

Figure 1. Example screen from the equation-balancing tutor.

questions if they are not made available for examination. For the struggling student, this type of help is critical to building a solid foundation. This is simply not practical in a non-interactive medium (textbook) or a non-intelligent software format like CAI. For further details on the design of the intelligent tutoring program and examples of student–tutor dialogues, please see reference 3.

Theoretical Framework

Students often have difficulty constructing effective mental models and representations in chemistry, especially concerning abstract unobservable concepts such as atoms and molecules. When students are unable to formulate these understandings, they suffer significant learning impediments [8] that limit their ability to write balanced equations, to comprehend the purpose of balancing equations, to interpret the symbolic representations used, and to solve problems based on equations. Clearly an effective tutoring system must go beyond traditional instructional models designed to place fully formed knowledge in the learner's path, to find ways to scaffold understanding, model problem-solving strategies, and increase students' ability to deal effectively with the task at hand. The design of the tutor uses natural language dialogue to support student learning in each of these areas.

An effective tutor must encourage learners to construct scientifically valid interpretations for the equation-balancing process while guiding them in altering their scientific misconceptions [9]. Contemporary research suggests that the social constructivist perspective—an approach that emphasizes the social contexts of scientific knowledge and that views learning as a collaborative, socially interactive, and social cultural activity [10]—provides a robust approach for developing student thinking and understanding in science. The equation-balancing tutor provides extensive opportunities for the student to "learn with" the tutor in co-constructing knowledge.

One of the most powerful ways that the tutor embodies social constructivist research is through its use of natural language to scaffold student thinking and performance while enabling the tutoring experiences to be embedded in learning situations that are as realistic as possible. During scaffolding, a more skilled individual (in this case the tutor) adjusts the amount of guidance needed to fit the student's current performance level. In a very real and intimate way, the tutor provides questioning, modeling, illustration, and explanation that grows more complex as the learner becomes more competent by adopting a specific form of scaffolding known as cognitive apprenticeship [11].

In cognitive apprenticeship an expert stretches and supports a novice's understanding and use of skills [12]. The term apprenticeship underscores the importance of modeling content-specific strategies for students. The tutor uses natural language to negotiate meaning and understanding by adjusting explanations and models of solutions to work within the student's range of understanding and to tap into the student's learning potential. Yet, it is not enough to scaffold students' ability to understand and formulate solutions. Like the very best human teachers, the tutor uses natural language questioning strategies in a second important way: to teach students how to ask good, thought-provoking questions about balancing chemical equations. Educational researchers argue that the ability to ask good questions might be the most important aspect of intelligence [13]. Recognizing that many students will not have the content or process language to ask effective questions, the tutor models effective questions to ask, prompting the student towards productive directions of thought and adding to their ability to use scientific language to explain and question their decisions and actions. Sometimes, when confronted with new concepts a student will be unable to formulate *any* meaningful questions, and getting examples of good questions is of tremendous benefit.

This approach provides students with two important sources of self-efficacy: confidence in their knowledge of the content of balancing equations and confidence in their ability to ask good questions about the process of balancing equations. Self-efficacy, the belief that one can be successful at the task at hand, has been shown to be an important factor in student motivation and goal setting [14]. Students who have a high sense of efficacy in a given area will set higher goals, be less afraid of failure, and adopt new strategies to replace those that fail. If efficacy for a specific task is low, students may give up easily, believing that they do not have the personal knowledge and skills to succeed. The tutor contains a built-in mechanism for building self-efficacy as students progress, as they become able to answer the questions for themselves and confirm their answers using the tutor.

Study Design

Students in CHEM 121 in the Fall 2001 semester at Duquesne University (N = 273) participated in the study. This is the first-semester general chemistry course for science majors. Lecture for the course was held three hours per week, with enrollment divided approximately 80%/20% between two faculty instructors. An additional recitation section was held one hour per week. There were twelve recitation sections and six recitation instructors. Six sections (N = 132) were

 S1430-4171(02)06620-7, Published on Web 10/24/2002, 10.1007/s00897020620a, 760379bj.pdf

Table 1. Student Performance on Balancing Equations (BE) and Entire Stoichiometry Quiz, Segmented by Total Score (See Text for Explanation of Categories)

	Group 1 Score < 20.2 N = 115	**Group 2** Score ≤ 18 N = 75	**Group 3** Score ≤ 17 N = 57	**Group 4** Score ≤ 16 N = 45	**Group 5** Score ≤ 15 N = 27
Test					
Quiz average (%)	61.0	53.9	51.2	48.2	41.8
BE average (%)	74.6	68.8	65.7	64.1	58.8
% total from BE	51.0	53.4	54.0	55.9	59.3
	N = 64	N = 44	N = 38	N = 32	N = 21
Control					
Quiz average (%)	69.6	64.8	60.7	57.7	45.3
BE average (%)	77.2	72.6	66.1	57.7	35.0
% total from BE	44.1	44.3	42.9	39.6	28.5
	N = 51	N = 31	N = 19	N = 13	N = 6

designated test groups and the remaining six sections (N = 141) were used as control groups. Each recitation instructor had one test and one control section, except for one instructor who had two test sections and one control section and one instructor who had a single control section.

After balancing equations was covered in lecture, in recitation students were given a homework assignment on balancing equations prepared by the course instructor. At that time, students in the test group also received instruction from the recitation leaders on how to use the equation-balancing tutor, and were instructed to use the tutor as part of their study in completing the homework assignment. Students used the tutor in a departmental computer laboratory and were required to submit a log file generated by the tutor as verification of completion of the assignment.

Quizzes were given weekly in the recitation section. On the week following the homework assignment, the quiz covered balancing equations as well as other topics in stoichiometry covered that week. The balancing equations portion consisted of five equations of varying difficulty, and was worth 10 out of 25 total points on the quiz. Students were directed to show as much work as possible to obtain partial credit. The results were analyzed as described in the following section.

Results and Discussion

On the balancing equations (BE) portion, the most frequent score in either group was 10 points (100%) and the average in both groups was well above 80%. With scores this high, it is difficult to extract the effect of the tutor at the upper end of performance. There was a high proportion of good students in the course; many had AP chemistry in high school and were already proficient in balancing equations. To assess the impact on the lower-performing students, for which this type of cognitive mentoring tutor is designed, the quiz results were segmented into groups according to total score. The first group consists of all students who scored below the class average on the quiz, which was 20.2 points (80.8%). To segment further according to low performance, successive categories consist of students with lower and lower total scores, less than or equal to 18, 17, 16 and 15 points, respectively. For each group, averages of three measures are presented: percentage on entire quiz, percentage on the BE portion only, and percentage of student's total score that came from BE. These results are given in Table 1.

These data provide several ways to view the impact of the tutor. The first and most obvious is direct comparison of test and control group scores on BE. For Group 1, the control group actually outperforms the test group slightly (77.2% to 74.6%); however, the control group students were consistently better performers than the test group in the course overall, as gauged by averages on the three quizzes given previously in the course (77.1% test, 83.4% control) and the first two hour exams (68.7% test, 74.3% control). This tendency is also reflected here in the total quiz scores, where the control group scores are on the order of 10% higher in all segmented groups except the very lowest-performing one (Group 5). On balancing equations, though, not only is the gap between test and control considerably smaller in Group 1 (2.6%), by Group 3 the gap is closed altogether, and in Groups 4 and 5 test outperforms control substantially, with the test group average higher by 23.8% in Group 5. There is a strong trend of improvement of test relative to control on BE in successively lower-performing groups, even though the control group remained consistently ahead on the quiz as a whole.

Next, we find that students in the test group consistently scored around 15% better on BE than on the quiz as a whole across all five scoring groups. This did not happen for the control group however. Though students also started out better on BE in Group 1 (7.6%), as we progress through the groups this declines, such that the untutored control group actually fell behind the quiz average by 10.3% by Group 5. By contrast, in Group 5 the tutored students' scores on BE were ahead of that group's quiz average by 17.0%. The trend in the test group is that the BE scores do not drop as fast as for the quiz overall, but they actually drop *faster* than the rest of the quiz for the control group.

It is also illuminating to examine the percentage of the students' total quiz score that came from their performance on the BE portion, reported in the third row for the respective test and control groups in Table 1. As total performance declines, for the test group the percentage of total points earned from BE sustains and steadily increases, from 51.0% to 59.3% from Group 1 to Group 5. One obvious explanation is the effect of the tutor, but an alternate explanation could be that students simply found the BE portion easier than the rest of the quiz, and so as total scores get lower the percent contribution from BE would naturally increase. If this were the case then one would also expect to see the same trend in the control group, but in fact the opposite is observed. For the control group, the percentage of the total score due to BE not only starts out lower than the test group in Group 1, it declines to only 28.5% by Group 5, compared to 59.3% for the Group 5 test students.

The issue of difference in time spent studying BE by the test and control groups should be addressed. Specifically, could the improvement have resulted simply because the test group spent more overall time studying BE than the control group, instead of being directly attributable to the tutor itself? This is not likely; the key is that no specific requirement on the amount of time to be spent using the tutor was mandated for the test-group students. They were simply instructed to use the tutor while studying balancing equations, and the amount of time actually devoted to studying with the tutor was up to each individual student. Therefore, because the test group students were not required to study more than the control group students, if they did in fact spend more time on balancing equations than they otherwise would, it is likely because they viewed the tutor as helpful or found it engaging. We expect there is a combination of factors at work in producing the improvement observed, including students receiving quality individualized tutoring as well as a motivational effect of the tutor causing students to want to spend more time studying. While it was not attempted to separate the effects of these individual factors, both are positive contributions from the tutor.

The clear differences in the various trends in test and control group performance all indicate that the tutor indeed improved student ability to balance equations, and that the extent of the tutor's support is significant, not marginal. For the students having the greatest difficulty, that is, as total scores decline, the impact of the tutor is the greatest.

While the trends described above are obvious from inspection, a further warrant of the evidence was sought via a χ^2 (chi-square) analysis to determine statistical significance. Because the student groups displayed in Table 1 are neither independent samples nor do they represent repeated measures of dependent samples, a test of significance differences across groups would be confounded by the problem of joint group membership; therefore, we analyzed the findings from only one group. Moreover, because the descriptive results supported the claim that the tutor is most effective for students who experience the greatest difficulty, the most likely candidate for significance testing would be the lowest-performing group, Group 5; however, the cell sizes in Group 5 were too small (only 6 students in the control condition) to provide an appropriate test of statistical significance. Group 4, the next lowest-performing group, was of sufficient size and thus afforded the best opportunity to compare expected performance with actual performance.

The best available predictor of expected performance on BE was the overall performance on the quiz. Overall quiz performance was used as the expected value and performance on the BE portion was used as the actual value in the χ^2 analysis reported here. For the control students in Group 4, there was no significant difference between expected performance and their actual performance on BE; however, for the test students in Group 4, there was a statistically significant difference between performance expected on the basis of the overall quiz scores and the students' actual performance on balancing equations, $\chi^2(1, N = 32) = 5.24$, $p < 0.025$. What is clear from the statistical analysis then, is that while students who did not use the tutor scored no better on balancing equations than would be expected based on overall quiz performance, the students who used the tutor did perform better than expected to a statistically significant degree. Therefore, both the descriptive analyses of the frequency groups and the inferential test for statistical significance support the positive impact of the tutor on student learning and performance.

Another question that can be raised is the possibility for contamination of the results through students in the control group finding out about and gaining access to the tutor from friends in the test group. This was a typical large first-year lecture course, in which most students would not know each other, therefore, although such contamination was possible, it was not probable to a large extent, and was expected to be minimized sufficiently by assigning different recitation sections to the test and control groups. The strong statistical significance of the improvement of test relative to control as indicated by the value of p above argues against any significant contamination, which would tend to weaken the correlation of test group membership with improvement on balancing equations. Nonetheless, it is worth noting that to the extent such contamination may have occurred, its effect on the results would be to dilute the observed improvement of test relative to control, in which case the impact of the tutor on student performance would be even more favorable than shown in Table 1.

Conclusion

This study establishes the feasibility of the AI-based approach to creating advanced new tutoring software for chemistry problem solving. In this case, the tutor improved the performance of the weakest students the most, as designed. The quantitative impact on the test group was shown to be statistically significant. Based on these results, we are excited about the continued investigation and development of these techniques, including extension to direct analysis of student work and development of tutors for additional chemistry topics.

The capability for students to ask questions of the tutor is particularly promising. This new cognitive modeling functionality allows interactive inquiry at a level not previously attained in chemistry software tutorials, directly supporting the goal of better teaching and learning of chemistry. The students characterized this feature as extremely helpful.

Further assessment of the tutor will be done with high school chemistry students. Here, the impact is expected to be even greater because these students are encountering balancing equations for the first time, while most of the students in the present study already had high school chemistry. The scope of assessment will be expanded to include qualitative evaluations such as student interviews.

Access to a Web-based demonstration of the equation-balancing tutor may be obtained by writing to us [15]. We are

 S1430-4171(02)06620-7, Published on Web 10/24/2002, 10.1007/s00897020620a, 760379bj.pdf

very interested in constructive feedback from readers of *The Chemical Educator*.

Acknowledgment. This work was supported by the National Science Foundation through Small Business Innovation Research award number DMI-0132003. We are indebted to the referees for helpful comments on the paper.

References and Notes

1. Lower, S.; Gerhold, G.; Smith, S. G.; Johnson, K. J.; Moore, J. W. *J. Chem. Educ.* **1979,** *56,* 219–227.
2. Bell, M. A.; Gladwin, R. P; Drury, T. A. *J. Chem. Educ.* **1998**, *75*, 781–785.
3. Johnson, B. G.; Holder, D. A. *Chem. Educator* **2002**, *7*, 297–302.
4. Herndon, W. C. *J. Chem. Educ.* **1997,** *74,* 1359–1362.
5. Windschitl, M.; Andre, T. *J. Res. Sci. Teach.* **1998,** *35,* 145–160.
6. Chi, M. T. H.; De Leeuw, N.; Chiu, M.; Lavancher, C. *Cognit. Sci.* **1994,** *18,* 439–477.
7. Zoller, U. *J. Chem. Educ.* **1987,** *64,* 510–512.
8. Garnett, P. J.; Garnett, P. J.; Hackling, M. W. *Stud. Sci. Educ.* **1994,** *25,* 69–95.
9. Resnick, L. B.; Chi, M. T. H. In *Science for the Fun of It: A Guide to Informal Science Education*; Druger, M., Ed.; National Science Teachers Association: Washington, DC, 1988; pp 24–31.
10. Rogoff, B. In *Handbook of Child Psychology*, 5th ed.; Kuhn, D.; Siegler, R. S., Eds.; Wiley: New York, 1998; Vol 2, pp 679–744.
11. Collins, A. In *Educational Values and Cognitive Instruction: Implications for Reform*; Idol, L.; Jones, B. F., Eds.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1991; pp 121–138.
12. Rogoff, B. *Apprenticeship in Thinking: Cognitive Development in Social Context*; Oxford University Press: New York, 1990.
13. Getzels, J. W.; Csikszentmihalyi, M. *Sci. J.* **1967,** *3,* 80–84.
14. Bandura, A. *Self-Efficacy: The Exercise of Control*; W. H. Freeman and Company: New York, 1997.
15. For more information, please email a request to johnson@quantumsimulations.com.