Intelligent Tutoring System Shows Effectiveness

Critique Of "Evaluating the Effectiveness of a Cognitive Tutor for Fundamental Physics Concepts" (Albacete, P. & VanLehn, K., 2000)

Introduction

Intelligent Tutoring Systems (ITSs) are different from other educational software because they have a sound theoretical background, and, their designs are driven by the theory. ITS applications have, to great extent, been designed and evaluated in lab environments with ideal assumptions set.

"Evaluating the Effectiveness of a Cognitive Tutor for Fundamental Physics Concepts" is a research paper in the 22nd Annual Meeting of the Cognitive Science Society. In this article, the author described and analyzed the evaluation of the Conceptual Helper, an ITS that uses cognitive approach to teach qualitative physics. The research study was carried out in real classroom settings. The results of the research showed a positive effect in gaining qualitative physics knowledge.

In this paper, I provided an introduction of the development of ITS theory, a summary of the article, reasons why I chose this article, and critique on the author's conclusions and suggestions.

Summary of the paper

Background

Conceptual Helper is an Intelligent Tutoring System (ITS) designed to coach students through physics homework problem solving of a qualitative nature. The design of Conceptual Helper is based on cognitive theories, specifically, the ACT theory of human cognition (Anderson, Boyle, Corbett, & Lewis, 1990).

Cognitive theory characterizes the structure of knowledge as semantic networks. Each proposition in the networks does not stand alone, but is linked to one another. Links are essential for reasoning qualitatively in solving physics problems. (For example, the link between acceleration and net force can be inferred from Newton's second law.) Several studies (e.g. VanHeuvelen, 1991) found that students' semantic network of conceptual physics is a collection of ill-structured, unconnected facts and concepts. In contrast, experts' knowledge is well structured and highly connected (Chi & Koeske, 1983). Based on these findings, the objective of Conceptual Helper is to help students establish and reinforce the correct links among physics concepts as they exist in the expert model, as well as to teach these concepts themselves.

Similar to other ACT (Anderson, et al. 1990) based tutoring systems, Conceptual Helper's main features include model tracing, user's on/off path control and immediate and individualized feedback. Conceptual Helper contains a cognitive model, which represents the knowledge that an ideal student (an expert) would possess about a particular subject. The tutor uses a process called model tracing. That is, every action performed by the student—entering a value, clicking a button, selecting a menu item, is checked against the cognitive model. Effectively acting in the same way that a human tutor would, if the student action is inappropriate, the tutor will not let the student go ahead. Instead, it will guide the student to a correct solution path. This matching is used as the basis for providing immediate feedback to students as they progress through the problem. The system also has a student model in which each node in the network represents a piece of conceptual knowledge that the student is expected to learn or a misconception that the tutor can help remedy. Each node has a number attached to it that indicates the probability that the student will apply the piece of knowledge when it is applicable. As the student solves a problem, the probabilities are updated according to the actions taken by the student.

Research Context and Methodology

In this research study, 42 students taking Introductory Mechanics classes were recruited and randomly divided into a Control group and an Experimental group. The length of the treatment was 2 hours.

1. Pre-test and post-test performance

Both groups took a pre-test that consisted of 29 qualitative physics problems. Then both groups solved some problems. The students in the Control Group had their input turned green or red depending on the correctness of the entry. Then, in the case of an incorrect action, the students could ask for help by making a choice from a help menu.

The kind of help they received consisted of simple hints such as “the direction of the vector is incorrect.” If the student asked for more help, they would just be told the correct answer. On the other hand the students in the experimental group received the green/red feedback depending on whether their action was correct but when the input was incorrect the Conceptual Helper intervened as explained above. After the students finished solving the problems with the system they took a post-test which was the same as the pre-test with the exception of a few changes in the cover stories of some problems.

2. Attitude questionnaire

The experimental group were asked to complete a questionnaire expressing their evaluation of the system.

Findings

1. Pre-test and post-test performance

A t-test of difference between pre-test scores and post-test scores is used to identify the effectiveness of Concept Helper. A significant difference was found ($t(40)=2.094$, $p=0.043$, two-tailed), which suggests that the intervention of the Conceptual Helper had a positive impact on the students’ understanding of the concepts as well as on their ability to abandon common misconceptions. In order to compare the effect of Concept Helper and that of other tutoring systems, an effect size is calculated by subtracting the mean of the gain scores of the control group from the mean of the gain scores of the experimental group, and divide by the standard deviation of the gain scores of the control condition. (Bloom, 1984) That calculation yields 0.63. Comparing with some other research results on other tutoring systems (e.g. Cohen et al., 1982; Bloom, 1984), this effect size, achieved within a two-hour period, is large enough to show the effectiveness of Concept Helper in terms of gains of scores. Other statistical data also show that the instructional intervention in the experimental group is more interactive than the control group, students with lower pre-test scores achieve higher gains in scores than students with higher pre-test scores (the experimental group), and a positive relationship between correction of misconceptions and coaching they received from the system.

2. Attitude questionnaire

The rating of the different aspects of the Concept Helper system was collected on a scale ranging from 1 to 5 where 5 was the best possible score. Students gave a score of 4 or above to all different aspects of the system (e.g., explanations that are clear to understand) which show a favorable acceptance of the system as well as a fairly high degree of liking of the mini-lessons (individual coaching units corresponding to each physics rule embedded in the questions).

Author's Conclusion

The evaluation of the tutor suggests that the teaching strategy followed by the Conceptual Helper along with its methodology for deploying the target knowledge and handling misconceptions, is effective in accomplishing the task it was designed to perform. The experimental group surpassed the control group in every statistical test performed. Moreover, a detailed examination of the effectiveness of each individual mini-lesson showed a trend in favor of using the lesson (though the small sample size could not provide significant evidence). The author also suggested that the reason of Conceptual Helper's effectives was the use of simple confrontation (as opposed to more elaborate and time consuming learning context, like experiments) and conceptual problems (as opposed to quantitative problems).

Reasons Why I Choose This Paper

Personally, I am very interested in Intelligent Tutoring System. It is a perfect combination of computer science and cognitive science. When tried Ms. Lindquist in the class, though a primitive form it was, its "intelligence" (tracing of cognitive model) surprised me by its effectiveness in coaching students problems solving skills in elementary algebra.

As we discussed in class, Intelligent Tutoring Systems are computer programs that are designed to incorporate techniques from the AI community in order to provide tutors which know what they teach, who they teach and how to teach it. AI attempts to produce a behavior in computer, which if performed by a human would be described as "intelligent". A current ITS should have four components: expert knowledge module, student knowledge module, tutoring module, and user interface. On the opposite of constructivist theory which stresses “learning by doing” or “learning by design”, ITS designers stress “learning by being told.” Due to this nature of ITS, it shows certain amazing advantages such as: clear articulation of knowledge, providing clear diagnosis of errors of students, showing how and why certain instructional techniques work or not. These features enable ITS to be a perfect test-bed for many theories and an ideal alternative of human tutors offering one-to-one tutoring, which was proven to be a very efficient way to attain higher achievements. In one research, 98% of students with private tutors performed better than those without private tutors. (Bloom, 1984)

From our reading assignment, we learned the overall characteristic of ITS, the development of a certain web-based ITS, and the mechanism of cognitive model tracing. We also learned from Anderson that most ACT systems were designed and evaluated in ideal lab environments and assumptions (Anderson, Corbett, Koedinger & Pelletier, 1995). The effectiveness in variable yet real classroom settings is unpredictable. After reading the assignments and exploring the Miss. Lindquist in class, I had following questions about ITS:

1. Does Miss. Lindquist significantly increase elementary school students’ achievement in algebra?

2. Does any characters of students affect their gains of achievement by using ITS?

a. Gender

b. Pre-treatment achievement in algebra

c. Pre-treatment computer skills

d. Interest and comfort of using computer

3. How do students like this software?

The paper I chose might complement our readings in that it focused on the evaluation of a certain ITS in real classroom settings and provided answers to some of the above questions. Since ITS is a novel branch of software and not many researches have been carried out, the research result in this paper is encouraging for future development of ITS and the evaluation methodology employed in this study may be replicated to other ITS evaluation research.

Critique of the paper

This research paper is based on real data in real classroom settings. The paper provides a very concrete methodology in collecting and analyzing data. In the beginning, the author briefly discussed the theoretical basis of ITS and the Conceptual Helper. Then the author described the detailed process of data collecting and analyzing. In this part the author also provided some previous research results on other ITS as benchmarks to show the effectiveness of Conceptual Helper. Finally the author discussed the possible reasons underlying the significant statistical results.

However, I come up with several issues that may improve this research paper that the author might have ignored:

1. In this research, the researcher split the subjects into 2 groups. The experimental group used the ITS while the control group used a manual help system to receive instruction (just same as using a reference book and looking for information by themselves). Hence, the statistical result only provided evidence that those students using ITS gained higher scores than those using a reference book. In other words, the ITS is effective as opposed to a reference book. This result is not sufficient to show how “intelligent” it is, when comparing with a human tutor. Since the cognitive model of ITS derives itself from human tutors, it is necessary to carry out a comparison between the two. I suggest that a third group be added into the research who receives tutoring from human tutors. If no significant difference is found between the score gains of the ITS group and that of the human tutoring group, the ITS will be considered an competent alternative of human tutors.

2. The sample size of this research is 42 students, and subjects were not recruited randomly (though they were grouped randomly). The sampling of this research might be biased. The size of this research is relatively small. When analyzing the data of the effectiveness of individual mini-lessons (tutoring units), the sample size was even not large enough to generate a statistical result. These factors seriously limit the research result’s ability of generalization to other students.

3. The research via the attitude questionnaire is not described in detail in this paper.

The author just concluded briefly that students liked this software. This information is not enough because some students' attitude or characteristic may affect their performance in taking tests, such as Pre-treatment computer skills, Interest and comfort of using computer, or even gender difference. Some other qualitative research methodology might be employed, i.e. observation, to collect information on students' attitudes, how they interacted with the computer and what difficulties they were encountered.

4. The length of this research is only 2 hours. Students' gains of scores or their fondness of this software might be, I believe, affected by the feeling of novelty. If the treatment were as long as a semester, would these students still like it and achieve significant gains of performance?

5. In the final discussion of the paper, the author attributed the success of Conceptual Helper to the strategy of "simple confrontation" while depreciated discovery learning or elaborate classroom settings like experiments. Students should be told "scientifically correct concept" directly. However, I don't fully agree with the author. Though the "simple confrontation" strategy shows its effectiveness in lower stages of learning (knowledge acquisition, etc.), it may not be effective when learners reach the level of expertise which requires learning from one's own experience. Constructivist theorists don't think so. From a constructivist viewpoint, student should learn by building his/her experience from doing authentic tasks. Although discovery learning

or elaborate physics experiments are complex, time consuming, or even not reflect on certain test measures, students can better incorporate the knowledge into their own experience, improve their meta-cognitive skills and apply the knowledge in the real world. The CoMPASS software which also helps students learn physics concepts, as we reviewed in class, showed positive and encouraging results in achieving its goal.

Conclusion

Instructional software, like all other educational material, should be evaluated before it is used in the classroom or research laboratory. (Heller,1991) This paper discussed the "Conceptual Helper" and its evaluation process and results, and proved that it successfully achieved its goal in improving students' performance in certain test measurements. However, during the evaluation, the whole classroom settings need to be evaluated via quantitative and qualitative study in stead of observing individual parts of it (i.e. achievement scores) while ignoring some other factors (i.e. students' attitudes, classroom, instructor, other resources, etc.). Although ITS shows its effectiveness in gains of scores of certain test measurements, other software based on different theoretical approaches may achieve different goals.

Reference:

Albacete, P. & VanLehn, K. (2000). Evaluating the Effectiveness of a Cognitive Tutor for Fundamental Physics Concepts. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, Erlbaum Associates: Mahwah

Anderson, J., Boyle, F., Corbett, T., & Lewis, M. (1990). Cognitive Modeling and Intelligent Tutoring. *Artificial Intelligence*, 42, 7-49. Elsevier Science Publishers B.V.

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, 4, 167-207.

Bloom, BS, (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring, *Educational Researcher13 4-16.*

Chi, M.T.H. & Koeske, R.D. (1983). Network Representation of a Child's Dinosaur Knowledge. *Developmental Psychology* 19(1), 29-39.

Cohen, P.A., Kulik, J.A., & Kulik, C.C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.

Heller, R. S. (1991). Evaluating software: A review of the options, *Computers & Education, 17 (4), 285-291*

VanHeuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of Physics*, 59(10), 891-896.