

CyberTutor Project Description

Introduction

This year we have made remarkable progress in using CyberTutor¹ to increase student learning in introductory calculus-based Newtonian mechanics, and to begin to study the learning process itself. In particular, we have shown that working more CyberTutor problems generates far more *improvement* in score on the final examination than does working more written homework problems (or going to recitation class more). Our research has revealed a surprising benefit to working a quantitative problem prior to related conceptual one, which may show a way to teach students to combine conceptual and quantitative reasoning in working problems. We have also shown that CyberTutor is an incredibly powerful assessment tool, with two orders of magnitude less error than a three hour final examination.

CyberTutor is a Socratic tutor for technical and scientific problems which attempts to reproduce the effect of a master teacher interacting one-on-one with each student. It allows us to author questions requiring free responses from the student, such as fill-in-the-blank, analytic formulae, simple graphs, etc. and to offer hints and simpler subproblems to the students at their request. Moreover, it responds usefully to over 50% of the wrong answers given, helping about 90% of the students through to the correct solution. We are using CyberTutor problems to replace about 60% of the written homework with an web-based problem solving experience that is more educationally effective, provides immediate feedback to the student and teacher, and produces a more reliable and detailed assessment of each student than has ever been possible before.

Study of Educational Effectiveness

CyberTutor is intended to be a replacement for written graded homework that teaches students more effectively. Therefore the most compelling testament to CyberTutor's educational effectiveness is that it has bested written graded homework in head to head comparisons of their relative ability to raise students' test scores. We studied MIT's introductory physics course² in Spring 2001 class because most of these students had taken an 8.01 final examination in the Fall, allowing us to study what factors caused a *gain* in the score on the final examination taken

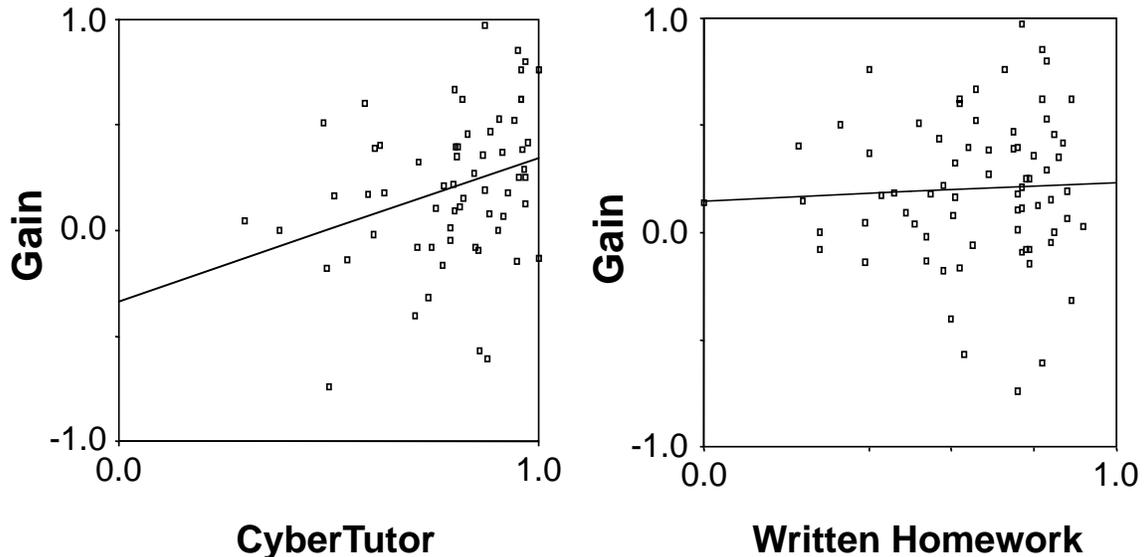


Figure 1. Normalized Gain on the final exam correlates ten times more strongly with CyberTutor than with written homework (data from MIT course 8.01, Spring 2001). Note: The CyberTutor grading algorithm results in scores that are highly correlated with how many CyberTutor problems a student completed.

the following spring. CyberTutor use was found to be the *only* course element that strongly correlated with gain (improvement) on the final exam (a combination of conceptual and quantitative problems). Fig. 1 shows the *gain* on the MIT final exam plotted against student's written homework and CyberTutor score. The correlation between CyberTutor score and gain on the final is ten times larger than for written homework, and 7 times higher than for participation in recitation class. This confirms that we have indeed found a better way than written homework to help students learn the material on our final examinations. In addition, CyberTutor proved comparable to group problem-solving in its correlation with gains on two standardized conceptual tests, outperforming written homework by a factor ~2, confirming the results by Ogilvie³ on the Spring 2000 course.

Student Opinion

Our CyberTutor problems are designed to be true “problems” rather than exercises which involve simply plugging numbers into formulas. They are almost exclusively free response, and 85% require symbolic solutions. Authoring, administering, and automatically grading such problems is challenging, but over the past three years we have learned how to write and assign problems of appropriate difficulty, and the CyberTutor grading engine, has been refined to the point that students feel they learn significantly more per unit time with CyberTutor than with written homework, as can be seen in Fig. 3. In addition they recommend that CyberTutor be used in future courses by a 5:1 ratio – a very high recommendation for any innovation (Figures 2,3).

We have felt it essential that CyberTutor use constitute part of the grade (typically 20%) so that the students perceive it as an important part of the course. However, this strongly diminishes student tolerance for perceived misgrading of one of their answers. Because CyberTutor is unique among web-based homework systems in accepting free form responses, it has been a particular challenge to ensure that student responses are fairly and consistently graded. Our student opinion results prove that both the program and our ability to use it has evolved to effectively meet this challenge.

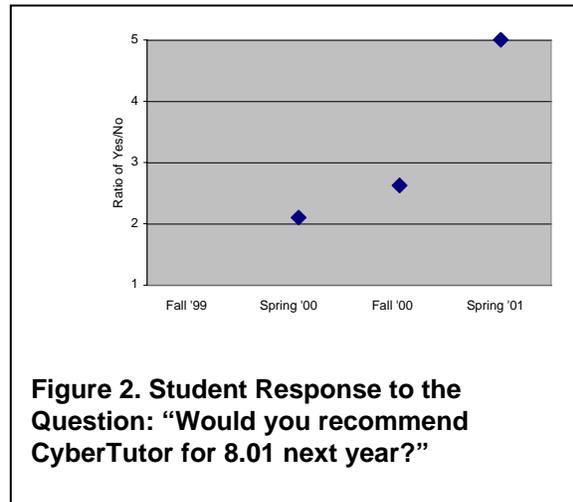


Figure 2. Student Response to the Question: “Would you recommend CyberTutor for 8.01 next year?”

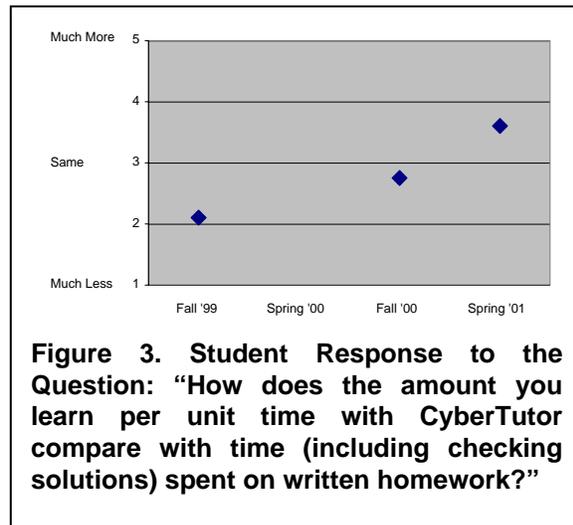


Figure 3. Student Response to the Question: “How does the amount you learn per unit time with CyberTutor compare with time (including checking solutions) spent on written homework?”

Assessment

Imagine that a rich ship-owner has hired Socrates to tutor his children. At the end of the month he desires to assess the amount they have learned. Would you advise him to:

- Administer a standardized hour-long test to the children?
- Ask Socrates how much they learned?

Of course, this question is rhetorical. Clearly Socrates' hours of individual interaction enables him to form a far more detailed and nuanced assessment than the test. In fact, Socrates probably maintains a continuously updated assessment based on *all* of his interactions with each student.

As would Socrates, CyberTutor integrates the traditionally separate activities of teaching and testing. Each CyberTutor interaction automatically and continuously generates data on student performance. Importantly, this highly detailed and reliable student assessment is available without the student stress and the lost class time of traditional testing. Moreover, assessments of CyberTutor data are available continuously and in real time.

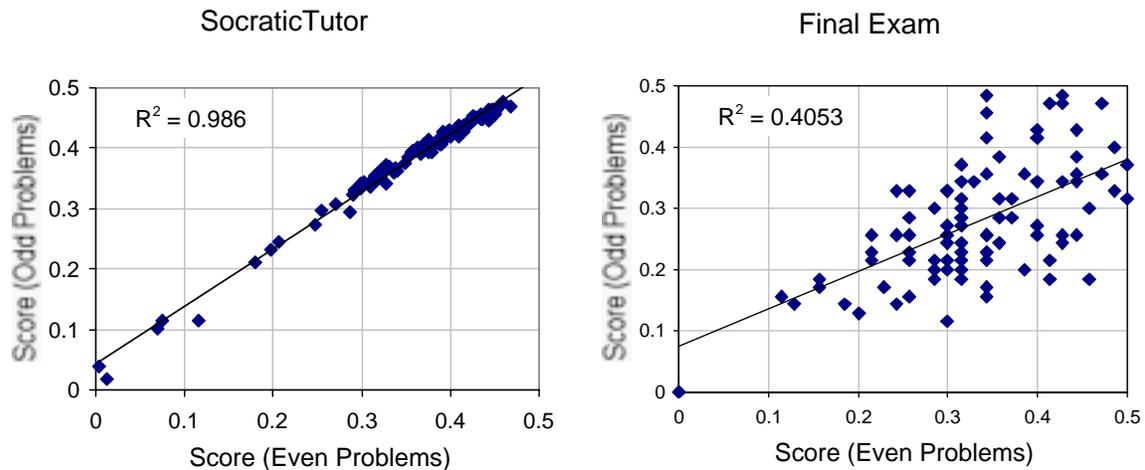


Figure 4. Total score on even versus odd problems in CyberTutor and on the final exam (data from MIT course 8.01, Spring 2001).

Statistical analysis of data from the Spring 2001 semester (Figure 4) shows that CyberTutor can assess students with about 100 times less error than a good three-hour final examination. Much of this huge improvement results from the fact that the typical student interacts with CyberTutor for about 17 times longer than the 3-hour test, diminishing the error due to lucky guesses and careless mistakes by roughly 17. Additionally, the CyberTutor assessment is about 6 times more reliable per unit time than the final exam. This additional reliability is obtained only after using information that is unavailable in a paper test—requests for hints/solutions, the number of wrong answers, and importantly, the time the student takes to complete each part of the problem. The bottom line is that when all this information is used optimally, a single CyberTutor problem assesses a student's skill nearly as well as a three hour final examination!

The CyberTutor analog of Socrates' highly detailed assessment is a **skill profile** – a skill rating on each of a set of predetermined topics. Having 100 times less error allows each of 100 skills to be determined with a reliability (freedom from error) equal to that of the final examination. The skill rating on a particular topic will be the best possible predictor of the relative class standing of that student on some established assessment vehicle on that topic. The topic might be a

13 - Physical Sciences – Cyber Tutor Project Description – 13 RLE Progress Report 144

conceptual topic (e.g. Newton's concept of force), a foundational skill like vector components, a general skill like using dimensions and units – or a regular syllabus topic like momentum.

The continuously updated CyberTutor skill profile provides an unprecedented level of feedback to inform the instructional process. It will enable revolutionary changes in education:

- Gaps in each student's foundational skills can be identified and remediated early in the course
- Students will know their state of knowledge well enough so that they can change their learning strategy from "avoid lost points" to "maximize learning"
- Accurate predictions of test scores can guide students studying for high stakes examinations
- Teachers can dispense with some testing and increase class time
- The skill level of the class on the current topic can guide the allotment of class time

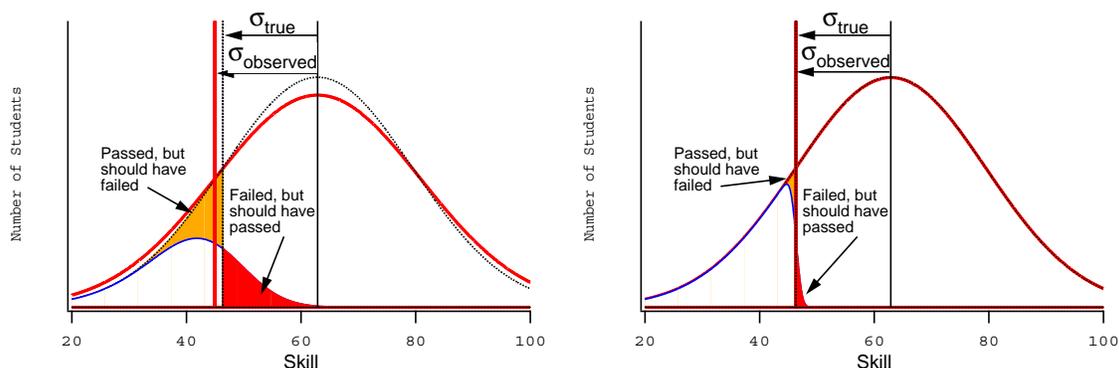


Figure 5. The dashed curve is the distribution of true skill and the thick curve is the distribution of observed skill (i.e. grade), widened by the random error. (Left) The 15% error variance on the final exam implies that for every 2 students who fail (i.e. score more than one standard deviation below the mean), one student actually deserved to pass or was passed but deserved to fail (Right) CyberTutor assessment has about 0.15% error. It would misdiagnose only 1% of the students, and these few would be only a point or two from the pass-fail line.

Educational Research

CyberTutor's detailed assessment will dramatically improve the quality of educational research, allowing researchers to obtain fine resolution on educational experiments in a short time. A typical protocol will be to divide students into two equivalent groups, provide each with a different educational item, then compare their results on a series of CyberTutor questions related to the item. This protocol is a unique and valuable tool for conducting physics education research and, with a class size ~100, is able to generate results valid at the 99% confidence level in a single trial.

For example, we recently measured the inductive influence between related conceptual and quantitative problems. (Something has a *positive* inductive influence on a subsequent assessment if it improves students' grades on that assessment.) While the conceptual problem didn't help students do a later quantitative one, as most physicists would generally suppose, working the quantitative problem helped significantly with a later conceptual one⁴.

Other educational research that could be performed using CyberTutor includes studying the inductive influence of problems requiring a numerical solution on subsequent problems requiring a symbolic solution, and studying which of two different explanations of a concept leads to a better student understanding of that concept.

Conclusion

Until recently, web based education has primarily meant bringing existing print materials verbatim to the web. CyberTutor represents the beginning of the next stage of web based education, when the interactivity of the web is utilized to full advantage. CyberTutor represents the leading edge in online homework delivery integrating educational effective Socratic dialog, immediate student and teacher feedback, and continuous, highly reliable assessment. This combination offers benefits to the entire educational community of students, teachers, and curriculum developers.

Students will learn more per unit time, and will enjoy the learning process. In the context of a course, students who have gaps in their foundational skills will be identified and helped early. Most importantly, CyberTutor will encourage students to develop their homework strategy around maximum learning rather than minimum lost points.

Teachers will save time creating, assigning, and grading homework. They will painlessly benefit from educational innovations that are built into CyberTutor. Informed by CyberTutor, teachers can adjust the allotment of class time to topics most difficult for their particular students. Finally, using CyberTutor assessment, teachers can decrease the amount of time devoted to examinations, and increase the amount of time devoted to teaching.

Curriculum developers will be able to use CyberTutor as a tool to refine their materials. They will be able to develop and deliver assignments that teach expert problem solving strategies, rather than novice exercises.

Finally, CyberTutor is a unique platform for educational research and its combination of flexibility, automated data collection, together with assessment of unprecedented reliability will enable research to uncover educationally effective innovations in educational pedagogy.

¹ CyberTutor is a product of Effective Educational Technologies, 88 Washington St., Cambridge, MA 02140.

² Pritchard, D.E. and Morote, S., "Effectiveness of 8.01 Course Elements in Spring 2000 and 2001" (2002). Available on the web: <http://torrseal.mit.edu/effedtech/pdf/smdp01.pdf>

³ Ogilvie, C., "Effectiveness of Different Course Components in Driving Gains in Conceptual Understanding" (2000). Available on the web: <http://torrseal.mit.edu/effedtech/pdf/ogilvie.pdf>

⁴ Dukes, P. and Pritchard, D.E., "Inductive Influence of Related Quantitative and Conceptual Problems" in Proceedings of the 2001 Physics Education Research Conference, Rochester, NY (2001). Also available on the web: <http://torrseal.mit.edu/effedtech/pdf/DukesPritchard.pdf>