What course elements correlate with improvement on tests in introductory Newtonian mechanics? In review Am J. Phys

Elsa-Sofia Morote and David E. Pritchard[a]
Physics Department,
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

In an MIT calculus-based introductory Newtonian mechanics course we study the effectiveness of various instructional course elements: electronic and written homework, collaborative group problems, and class participation. We measure effectiveness by the slope of the regression line between a student's score (used as a proxy for participation) on a particular course element and his normalized gain on the various assessment instruments. These instruments were the MIT final exam comprised mainly of multi-part problems demanding analytic responses, and two widely used standard physics tests that emphasize conceptual knowledge: the Force Concept Inventory and the Mechanics Baseline Test. The results show that interactive course elements are associated with higher gains on assessment instruments: doing interactive electronic homework administered by myCyberTutor correlated with large gains on the final exam producing a learning effect of 1.8 ± 0.4 standard deviations on the final examination score. MyCyberTutor and collaborative group problem solving correlated with gains on the more conceptual tests. We also report surveys that demonstrate that students have had an increasingly favorable opinion of myCyberTutor over the four terms of its use.

---

[a] Electronic address: dpritch@mit.edu

I. INTRODUCTION

The applied side of physics education research attempts to associate students' improved knowledge (often measured by gain between before and after testing) with some identifiable instructional element thereby to identify and/or improve elements that are effective at enhancing performance. In this spirit, our study measures correlations between before and after test gain and various course elements: participation in recitation sections and tutorials, and scores (used as proxies for participation) on group problems, interactive electronic homework, and conventional written homework.

The results are relevant to discussions of the impact of curriculum innovation (Van Aalst[1] in physics), and confirm the superiority of interactive elements (e.g. interactive lecture demonstrations[2] peer instruction[3] and group problem-solving[4,5]) for imparting conceptual learning. The present study is unique in that it extends this type of before and after study to performance on hand-graded multi-part final examinations. It is also unique in including a study of a highly interactive electronic homework tutor called myCyberTutor[6] which this study shows produces a learning gain of nearly two standard deviations on the final examination.

Interactive methods are those that require interactive engagement (based on Hake[7]'s definition) of students with "heads-on (always) and hands-on (usually) activities" that yield immediate feedback through peers, instructors, or intelligent computer programs. Traditional methods are those that "make little or no use of innovative methods, relying primarily on passive-student lectures, recipe labs, and algorithmic-problem exams."

Hake[7] has compared interactive-engagement versus traditional methods using pre- and post-test data obtained from the Force Concept Inventory (FCI) and Mechanics Baseline test (MBT). These tests are complementary probes for measuring understanding of basic Newtonian concepts. Questions on the FCI test[8] were designed to be meaningful to students without formal training in mechanics and targets their preconceptions on the subject. In contrast, the MBT[9] emphasizes concepts that cannot be grasped without formal knowledge of mechanics.

Hake[7] obtained data from both tests administered to 6,500 students in 62 courses, and showed that the average normalized gain (g) is a good metric for "course effectiveness in promoting conceptual understanding." The normalized gain is the improvement in score normalized by the maximum possible improvement; it is determined from the "postest" (S_{after}) "pretest" (S_{before}) examination scores:

$$g = \frac{S_{after} - S_{before}}{100 - S_{before}} = \frac{Actual\_gain}{Maximun\_possible\_gain} \quad (1)$$

Hake[7] found that classes that used interactive-engagement methods outperformed traditional classes by almost two standard deviations with respect to the normalized gain. He found that traditional classes had an average normalized gain equal to 0.23, whereas classes using interactive methods obtained an average gain of 0.48 ± 0.14 (std. dev).

In the same vein, utilizing the FCI test, Saul[10] compared student learning of mechanics in traditional (lecture and recitation) first-semester calculus-based physics courses with three innovative curricula: McDermott's *tutorials*[11], Heller's *group problem-solving*[4,5], and Law's *workshop physics*[12] . The curricula included lecture, lab, and recitation combined into three two-hour guided discovery lab sessions a week. As in Hake's[7] study, Saul[9] confirmed that traditional classes average about 0.20 normalized gains, and the innovative curricula (tutorials and group problem solving) average 0.37 gains, while guided-discovery instruction (workshop physics) averaged 0.43 for the normalized FCI gain (See Table I).

Ogilvie[13] used a method similar to Saul's[10] analysis but added an important course element: interactive electronic homework. He administered the FCI test to approximately 100 students before and after they took the Spring 8.01 class (Calculus-based "Introductory Newtonian Mechanics") at the Massachusetts Institute of Technology (MIT) in 2000. Ogilvie[13] then provided data on the correlation of various course elements such as *tutorial attendance*, *written problem sets,* Pritchard*'s interactive electronic homework (myCyberTutor)*[6] and *collaborative group problem-solving* with each student's gain on FCI test (see both Table I summarizes previous studies, and section II on course overview)

Homework, in general, has been appreciated as an effective course element. Cooper[14] found at least 50 studies that correlated the time students reported spending on homework with their <u>achievements</u> (not the <u>improvement</u>, as studied here). Cooper[14] affirmed that homework has several positive effects on achievement and learning, such as improved retention of actual knowledge, increased understanding, better critical thinking, and curriculum enrichment.

Electronic homework as a course element has more positive effects than written homework according to some researchers. Mestre et al.[15] compared the effect of electronic and written homework on student achievement by measuring exam performance. They found that electronic homework correlated with higher overall exam performance. Thoennessen and Harrison[16] confirmed that electronic homework has a clear correlation with the final exam score, and found that students prefer using it to written homework. The electronic homework tested by these researchers contains clear pedagogy and students received instant feedback and hints. The pedagogy implemented in the electronic homework is important. For instance, Bonham et al.[17] found that electronic homework systems with standard textbook-like problems with numerical answers and no informative feedback do not provide more significant benefits than written homework.

II. COURSE OVERVIEW

Calculus-based course 8.01 at MIT "Introductory Newtonian Mechanics" is one of the most difficult courses required of all MIT graduates. Typically, 15 % of entering freshmen fail to receive a grade of C or better and are therefore forced to repeat it. Consequently, more than 90 % of students taking 8.01 in the Spring term of this study had previously attempted this course; they had not learned how to solve the mostly multi-part problems requiring symbolic answers. In the Fall there are three small enrollment versions of 8.01 in addition to the "standard version." However, most Spring term students came from the "standard 8.01." We could not find any significant difference between these students and those from the smaller courses. This study reports data from Spring 8.01semesters in 2000, 2001, and 2002.

These Spring 8.01 courses that we studied had recently been re-organized to better teach relevant problem-solving skills. It did not use lectures to present new material. This was not a radical step since most of these students had the opportunity to attend lecture-demonstrations in their previous 8.01 courses. "New" material was introduced in the three recitations on Monday through Wednesday, reviewed in tutorials on Thursday, and reviewed and tested on Friday. Homework problems were required in two formats: in conventional written form and electronically, using myCyberTutor. Attendance and participation in recitations constituted 3% of the grade; a challenging group problem, counting for 7%, was given to groups of two or three students in class each week.

The Spring course utilized the following instructional course elements:

A. Interactive Methods

Interactive Electronic Homework

This was an electronic tutoring system, myCyberTutor[6]. It behaves like a Socratic tutor, presenting problems and offering students help upon request in the form of hints and simpler subtasks, and provides helpful suggestions or spontaneous warnings when particular incorrect answers are given. It tutors more than 90 % of the students to achieve the correct solution, charging a modest 3 % penalty for hints used; hence the myCyberTutor grade is primarily an indication of how many problems are attempted with it. Most problems have multiple parts that demand free response symbolic answers. About 15 % are conceptual questions, many motivated by Physics Education Research.

Group Problem Solving

Students worked in groups of two or three to collaboratively solve difficult (but not context-rich) problems in a manner as pioneered at the University of Minnesota[4,5].

B. Traditional Methods

Written Homework

Written homework contained mostly original problems written by the instructor David Pritchard; many involved real-world applications of physics principles (i.e. were more context rich) than standard end-of-chapter problems, and skills such as scaling and estimation were often involved. Solutions were provided on the due date and all problems were graded by hand.

Class Participation

In 2001 and 2002 only, students received participation grades in recitation sessions based on a weighting of attendance (67%) and participation in discussions (33%). There were three recitations / week in this course; plus a single half-hour review lecture and a half-hour test.

Tutorial Attendance

Small tutorials were required of all students in 2000, but were required only of under-performing students in 2001. Three or four students met with a senior undergraduate or graduate tutor (TA) for a one-hour tutorial in which everyone helped each other on typical weekly exam or homework problems. No special instructional material, training or guidelines was provided for the TAs. Tutorial attendance was measured for this study.

III. METHODOLOGY

The central question in applied educational research is "How does this learning activity affect the amount learned?", perhaps per unit of student time spent on the activity. We are interested in finding the average normalized gain (to measure amount learned) associated with each instructional element. Ideally this would involve a comparison of two classes, distinguished only by one extra course element that all students in the "experimental class" used exactly as intended by the instructor. Rather than implement this approach with a different physical class for each instructional element involved, we instead find virtual classes within one large class that differ by the amount of that course element that they elect to use. We contrast our "correlation" method with the more usual one of using two separate classes at the end of this section.

The mathematics behind our method is straightforward: we find the dependence of the student's normalized gains on each course element using linear regression; then compare the performance of students who use the average amount of that element with the gain of those who use none. We call this difference the Extrapolated Gain because the linear extrapolation of the normalized gain vs the score to zero score creates a control class - one that did not use that instructional element.

The first step is to use standard linear regression methods to fit the normalized gain vs. the score on an each instructional element to the expression: $g_i(s^i) = c + \beta^i \times s^i$

where s^i is the gain score on the ith instructional element and β^i is the slope for that element (and for the particular assessment used, i.e. either the MIT final exam or a conceptual test (e.g. FCI or MBT)). The Extrapolated Gain (G) for each course element on each assessment is then defined as

$$G = S_{avg}^{i} \times \beta^{i} \qquad (2)$$

where S_{avg}^{i} is the average score of the class on that particular course element. The right hand side of this equation can be thought of as $S_{avg}^{i} \times \beta^{i} - S_{avg}^{i} \times 0$ the difference in normalized gain between the average in our class and an extrapolated class that did not use that instructional element. Thus G represents the class' normalized gain improvement on that particular assessment that correlates with that course element. Normalized gains require before and after testing. The Force Concept Inventory (FCI) was administered before and after the 8.01 course in Spring 2000 by the instructor. The MBT, which contains a small fraction of numerical problems and also covers energy and momentum, was administered before and after the Spring 2001 course by the instructor. The normalized gain on the final examination ("posttest") was computed for those students who had taken a final exam in 8.01 ("pretest") at the end of the prior Fall semester.

Important to our use of score on the various course elements as a proxy for amount used is that the score primarily represents the amount of learning activity attempted, as opposed to the amount of skill achieved. Clearly, the recitation participation grade and tutorial attendance are pure instructional activities; scores indicate only that the students participated, not that they did well. Since more than 90% of students using myCyberTutor successfully completed each attempted problem (with very little penalty for hints), and since average scores on written homework were generally ~85% for those problems attempted, homework scores were primarily an indication of the number of problems and assignments attempted, and are therefore largely instructional (rather than an assessment). The group problem-solving was clearly part learning activity and part assessment, being a graded class exercise. However, students whose overall score was below average were almost always those who did not attend several of the group problem-solving sessions, those who attended faithfully generally received above average scores.

It is worth noting the differences between our correlations-based methodology and the more common one of giving class "E" (for experimental) one treatment and class "C" another (or none if it is a "control" class). This latter type of study is ideal for deciding which treatment is better, but it determines only the differential effect of the treatments (generally the control class engages in some other learning activity during the time that the experimental class undergoes the experimental treatment). In contrast, correlation shows a relationship to each (of possibly several) individual instructional element. Correlation can compare several different elements in one study, whereas the E versus C approach requires additional experimental classes when more than one instructional element is being studied. Both methods have potential pitfalls: when using E vs C care must be taken that no other factor is different between the two groups; in correlation studies, it is possible that some hidden causal factor creates the correlation (see discussion below). One drawback of the correlation approach is that it requires a larger sample to produce results of the same statistical validity as the E versus C approach if few students elect to do little of a particular course element (whereas all students in the control class do none). If normalized gain is used as the metric, both methods will have to cope with large scatter in the data (e.g., in Fig 1) because of the compounding of the random testing error when subtracting pre and post test scores to compute the normalized gain.

IV. GAIN ON THE MIT FINAL EXAM

The majority (70%) of the 8.01 course students in Springs 2001 and 2002 had taken a final examination (the pretest in our study) in one of the four versions of 8.01 during the previous semester with an unsatisfactory result. None of the Fall semester exams had a significant conceptual component. The class averages for the finals in all Fall versions of 8.01 final were similar and the finals were considered equivalent. (Attempts to cross calibrate these finals, e.g. on the basis of entering MBT scores, were unsuccessful, mostly because there were typically fewer than six students in each class.) Obviously, our entire sample took the Spring final as the posttest for finding the normalized gain on the MIT final.

The final exam in the Spring courses consisted of about ¼ conceptual questions because it includes the MBT. This is a significant deviation from the usual MIT consensus that only problem-solving questions are required for MIT students in 8.01 and in most engineering and science disciplines. The results in this paper used only the problem-solving grades to compute the gain, although including the MBT part of the final would change the extrapolated gain by considerably less than one error bar.

The correlation between each course element and the normalized gain on the final was found using linear regression from data like those shown in Figure 1. The straight regression lines show the relationship between normalized gain and Written Homework (left panel), and between normalized gain and myCyberTutor (right panel). The myCyberTutor slope implies that a student obtaining the average score on myCyberTutor would have an extrapolated gain of 0.55 (see Table II) relative to one who did not use myCyberTutor at all. The correlation with Written Homework is positive (as it also was in Ogilvie's study[13]), but small and statistically insignificant. Note that the slope, β does is invariant if the before and after finals have different average score because a difference would merely displace all points up or down. (There would be some effect on normalized gain if the standard deviations were different, but the final exam scores all had standard deviations in the 13.5% to 15.0% range which had negligible effect on this analysis.)

Fits to data in Figure 1 for 2001 are summarized in Table II, which shows the slope (β) of the regression line in the first column. Displayed in the next two columns are the extrapolated gain attributable to each element (from Equation 2), along with its standard error (δ_{Gain}), which is the standard error in β times the average score. The last column represents the p-value of β. Data for the 2002 class were similarly processed, and the results are presented in Table III.

All data for 2001 are presented, but data for two from five sections of the 2002 class are excluded in the similar Table III for 2002 because the professor in charge of those sections did not encourage his students to use myCyberTutor and prepared them specifically for the final examination. (A t-test detected that the performance of students in those sections was significantly different ($p < 0.05$) from the others.) The extrapolated gains and the p-values inferred from the improved statistics are presented in Table IV and also summarized in Figure 2. The extrapolated gains determined for the

2001 and 2002 classes were at or below the combined error for all course elements except group problem-solving where the difference was ~ twice the combined error. Since the group problem-solving element was similar in both years, this difference is attributed to statistical fluctuation (about 20% likelihood for this discrepancy to occur by chance in our study since it makes four comparisons).

With respect to before and after final exam scores in both years, myCyberTutor had the highest slope and was statistically significant ($p < 0.05$) in both years (see Table II & III). Written homework, group problem-solving, and class participation did not show significant correlation with the gain on the final (note the difference between the slopes of written homework and myCyberTutor in Figure 1) except that in 2002 group problem-solving was a significant contributor to the final exam's gain (Table III).

To place these results in educational context, a standard educational metric was used. The effect size, the improvement, was measured in standard deviations. This requires knowledge of the score improvement and standard deviation. The effect size[18] is simply (postscore – prescore)/(standard deviation):

$$d = \frac{S_{after} - S_{before}}{\sigma} \qquad (3)$$

$$d = \frac{G(1 - S_{before})}{\sigma} \qquad (4)$$

Equation 4 follows from Equation 1 for G and we take σ to be the standard deviation of the before scores. In this equation, values from the 8.01 final exam in the preceding semester are used. For 8.01 in the Fall 2000 the students who had to repeat the course (mostly in Spring 2001) averaged 45.0%. The standard deviation was σ =14.7% (so they averaged 1.7 standard deviations below average). Equation 4 then gives an effect size $d_{2001} = 2.14 \pm 0.82$. The corresponding numbers for the 8.01 Fall 2001 course were 44.7% and 15.1%, yielding $d_{2002} = 1.65 \pm 0.48$. These numbers average to $d_{2001\text{-}2}$ 1.79 ± 0.41.

Educational interventions are considered successful with an effect size of 1.0, and 2.0 is considered exceptional[18], so the correlation between doing the electronic homework and improvement on the MIT final exam is very encouraging. However, it must be borne in mind that the extrapolation to zero effort necessary to find the extrapolated gain using a linear fit (improvement assumed proportional to amount done), while reasonable, is not strongly tested with the data at hand due to the small number of students who did very little electronic homework.

V. THE GAIN ON FORCE CONCEPT INVENTORY AND MECHANICS BASELINE TESTS

The Force Concept Inventory was administered before and after the 8.01 Course taught by C. Ogilvie in Spring 2000. Scatter plots of normalized gain versus course elements are contained in Ogilvie[13]. Re-analysis of the data (taken from Ogilvie's graphs) are given in Table V and Figure 3. myCyberTutor and group problem-solving show the most significant Extrapolated Gains on the FCI.

In 2001 and 2002, Pritchard administered the Mechanics Baseline test before and after this course. A good measure of the class is the "before" grade on the MBT, within 0.2 of 13.5 (out of 26 graded with no penalty for wrong answers) each year[b].

Table VI shows individual regressions between each course element and the MBT normalized gain for 2001. Group problem-solving and myCyberTutor show the most significant extrapolated gains on the MBT (p-value < 0.06). Written homework showed a higher gain than group problem-solving, but it has high error and therefore a large p-value. Class participation shows no significant effect (Figure 4, Table VI). In 2001 one of the MBT problems was incorrectly graded (discovered only after the tests were destroyed). This could lower the gain by 0.06 at the maximum, less than the stated errors. The elimination of the two class sections in 2002 reduced all correlations with the gain on the MBT to below statistical significance.

In summary, significant Extrapolated Gains on the more conceptual tests (MBT and FCI) occur with both myCyberTutor and group problem-solving. The improvement on the conceptual tests due to the electronic homework might be termed encouraging because imparting conceptual knowledge was only a minor goal of the problem design and selection. The effect of group problem-solving is also encouraging. Although students spend much more time on myCyberTutor than on group problem-solving, most of it is spent on multi-part problems (the 15% conceptual questions consume significantly less than 15% of the time). Hence it is not clear whether group problem-solving or

---

[b] This is the same regular Fall 2003 version of 8.01, and is typical of highly selected university classes - e.g. the honors freshman class at Ohio State University: L. Bao, private communication.

conceptual questions in myCyberTutor correlate more strongly with extrapolated conceptual gain per unit of time on that task.

VI. STUDENT OPINION OF MY CYBERTUTOR

Student opinion was gathered concerning the educational effectiveness of myCyberTutor and the desirability of using it in 8.01 in the future. This provides complementary information about myCyberTutor's effectiveness, and about its overall level of student acceptance. It is important because no educational innovation is likely to be successful without student acceptance.

Two questions were generally asked of the students on the end-of-term questionnaires about myCyberTutor. One assessed myCyberTutor learning relative to written homework; the other addressed the desirability of continuing to use it. The strong upward trend of the data on the accompanying graphs indicates that continued use of myCyberTutor was highly recommended, most recently by a 7:1 ratio (Fig. 5, right panel). Students may feel that they learn significantly more per unit time when using myCyberTutor than when doing written homework (Fig. 5, left panel). This confirms Thoennessen and Harrison's[16] findings that students prefer electronic homework over written homework.

VII. DISCUSSION

This study shows once again that traditional methods of instruction do not yield large extrapolated gains on conceptual tests, and extends this conclusion to final examinations with multipart problems demanding symbolic answers. Attending tutorials or recitation did not yield significant extrapolated gains on the MIT final in any one case, although the gain was positive in all cases and the result of a weighted average would show a small but significant correlation between attendance and improved performance on the final exam. On the other hand, tutorials and class attendance had (negative but) insignificant extrapolated gains on both conceptual tests. This probably reflects that these venues mostly emphasize the algorithmic steps necessary to solve the particular problems on the previous weekly exam (tutorials) or on the current homework problems (recitations). Note that the slightly negative correlation of tutorials and gain could result because students who do problems collaboratively outside class (learning concepts in these discussions) do better on weekly tests and feel less attend tutorials and complete the homework so they don't feel a need to attend recitation on the day the homework is discussed.

The verdict on written homework is more positive; the correlation is positive in all four cases for which data were presented, generally with p-values between 0.1 and 0.2. Taken together, these indicate a marginal gain due to written homework on the final exam and a barely significant gain on the conceptual tests. The small but significant gain attributed to written homework may reflect the fact that it is the most interactive of the traditional instructional elements studied here.

The interactive instructional elements – group problem-solving and electronic homework – had the highest extrapolated gain in this study. Likely reasons for the success of group problem-solving are given in Anderson, Heller, Hollabaugh, and Keith[4,5]. MyCyberTutor showed a very strong correlation with gain on the MIT final, a learning effect of 1.8 is nearly what a personal human tutor could achieve (albeit perhaps with less student time). This is encouraging, as myCyberTutor's content was designed to increase skills tested on examinations. This suggests that if its content were designed to teach some other skill (e.g. estimation, checking your answer, Chemistry) it would do very well on this activity also. MyCyberTutor compares well with group problem-solving in correlations on the conceptual tests, a technique known to teach concepts effectively[4,5].

One possible explanation for myCyberTutor's effectiveness (especially relative to other online homework systems) is that it is an interactive tutor, not simply a homework administration system. It offers spontaneous feedback about particular wrong answers, several types of hints are available on request, and follow-up comments and follow-up questions make students ponder the significance of their solution before rushing on to the next assigned problem. Moreover, these features are heavily used – students make an average of ten round trips to the computer while working through each problem. This contrasts with student focus on solely obtaining the answer on written homework (as well as on electronic homework administration systems that respond only by grading answers right or wrong). A second advantage of electronic homework over written homework is that copying of the latter is endemic and has low instructional value. In contrast, student response patterns on myCyberTutor showed that only about 4% of all students had the conspicuous lack of wrong answers given and hints requested that strongly suggested that were obtaining many of "their" solutions elsewhere. This rate of unauthorized collaboration is a far lower rate of academic dishonesty than is reported on written homework on self-reported surveys of academic dishonesty at MIT and elsewhere[19].

It is tempting to dismiss these results as "just a correlation - perhaps the good students found myCyberTutor easier to use and used it more." Such arguments do not work since the correlations here are with improvement, rather than with score. The good students would have done better on the pretest as well as the posttest; thus being a good student does not by itself correlate with increased gain. There is however, a more subtle possibility for the observed correlation: myCyberTutor may appeal more to students who are learning more (e.g. because they get immediate positive feedback when they figure something out) and hence those students who are going to show the highest gains will be inclined to do more of it. Without further elaboration, this explanation does not address the observation that the gains on the MIT final correlate much more strongly with myCyberTutor use than do gains on the conceptual tests. The most straightforward explanation for the correlation is that students learn the test material (and receive scaffolding for problems like those on the MIT final) by using myCyberTutor. This would be expected since the myCyberTutor content was designed to help students with multi-part problems requiring symbolic answers.

In summary, four independent studies (derived from linear regression with gain on the MIT final, MBT, and FCI) show that student scores on interactive instructional elements such as interactive electronic homework and group problem-solving correlate more strongly with gain on assessment instruments - both conceptual and symbolic - than do scores on traditional elements. MyCyberTutor is far and away the most effective course element as judged by correlation with improving final examination scores. Group problem-solving is the second most effective course element, and correlates at least as well as interactive electronic homework with extrapolated gains on standard tests emphasizing conceptual knowledge. myCyberTutor has also received an increasingly favorable comparison with hand-graded written homework and enjoys a very strong (7:1) recommendation from the students that its use be continued in the future.

For the future, this study suggests that efforts to improve end-of-term test scores in “Introductory Mechanics” at MIT should concentrate on improving interactive instructional activities. Improving interactive electronic homework, especially for conceptual material, and finding recitation and tutorial formats that are more interactive would both seem to offer rewards.

Tables

Pre and Post Test	Research	"g" Traditional Methods	"g" Innovative Methods
MBT	Hake	0.23	0.48 (Interactive methods)
FCI	Saul	0.20	0.37 (McDermott's tutorials) 0.37 (Heller's group problem solving) 0.43 (Law's workshop physics)
FCI[c]	Ogilvie	0.14	0.30 (Heller's group problem solving) 0.39 (Pritchard's myCyberTutor)

TABLE I. Results of previous studies

[c] in this case the values are the extrapolated gains (see Table V)

Course Element	β	S_{avg}/S_{max}	G	δG	p-value
myCyberTutor	0.688	0.801	0.551	0.211	0.010
Written Homework	0.083	0.665	0.055	0.140	0.690
Group Problem Solving	0.056	0.624	0.035	0.090	0.690
Class Participation	0.116	0.621	0.072	0.075	0.345

TABLE II. 2001 Final exam improvement vs course element, N = 64 students. Savg and Smax are average and maximum scores respectively. G is the extrapolated gain and G its error.

Course Element	ß	S_{avg}/S_{max}	G	δG	p-value
myCyberTutor	0.769	0.89	0.411	0.120	0.003
Written Homework	0.702	0.89	0.385	0.245	0.131
Group Problem Solving	0.480	0.86	0.248	0.082	0.007
Class Participation	0.334	0.78	0.156	0.095	0.115
Tutorial Attendance	0.337	0.80	0.173	0.108	0.124

TABLE III. 2002 Final exam improvement vs course element, N = 38 students. Savg. and Smax. are average and maximum scores respectively. G is the extrapolated gain.

Course Element	G	δG	p-value
myCyberTutor	0.445	0.104	0.00002
Written Homework	0.136	0.122	0.26
Group Problem Solving	0.151	0.060	0.013
Class Participation	0.104	0.059	0.076

TABLE IV. Final Exam gain score versus Course Elements - combined years 2001 and 2002. Effect size $d_{2001\text{-}2} = 1.79 \pm 0.41$.

Course Elements	β	S_{avg}/S_{max}	G	δG	p-value
myCyberTutor	3.73	0.815	0.395	0.181	0.02
Written Homework	1.66	0.688	0.141	0.124	0.2
Group Problems Solving	3.87	0.782	0.301	0.198	0.09
Tutorial Attendance	-0.27	0.846	-0.020	0.121	0.854

TABLE V: Improvement on the Force Concept Inventory vs course element (2000-Based on Ogilvie[13], N=56). Savg and Smax are average and maximum scores respectively. The β values are higher here because Ogilvie used a different scale for the scores. The extrapolated gains, G, are comparable; students.

Course elements	β	S_{avg}/S_{max}	G	δ G	p-value
myCyberTutor	0.264	0.680	0.21	0.125	0.059
Written Homework	0.297	0.715	0.212	0.131	0.109
Group Problem Solving	0.267	0.795	0.181	0.066	0.008
Class Participation	-0.072	0.677	-0.049	0.057	0.39

TABLE VI. 2001 gain on the Mechanics Baseline Test versus course element, N=64 students.

Figure Captions

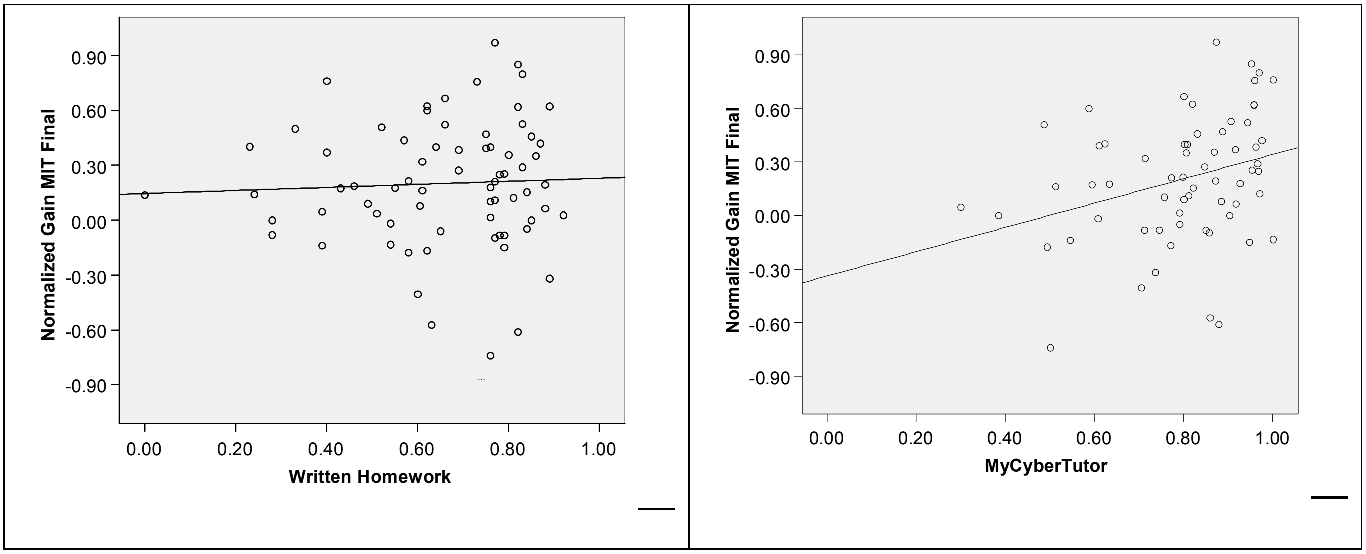

FIG. 1 : Normalized gain (g) on the final exam versus written homework (left) and vs myCyberTutor (right) – 2001. (Reprinted from ref. 20.)

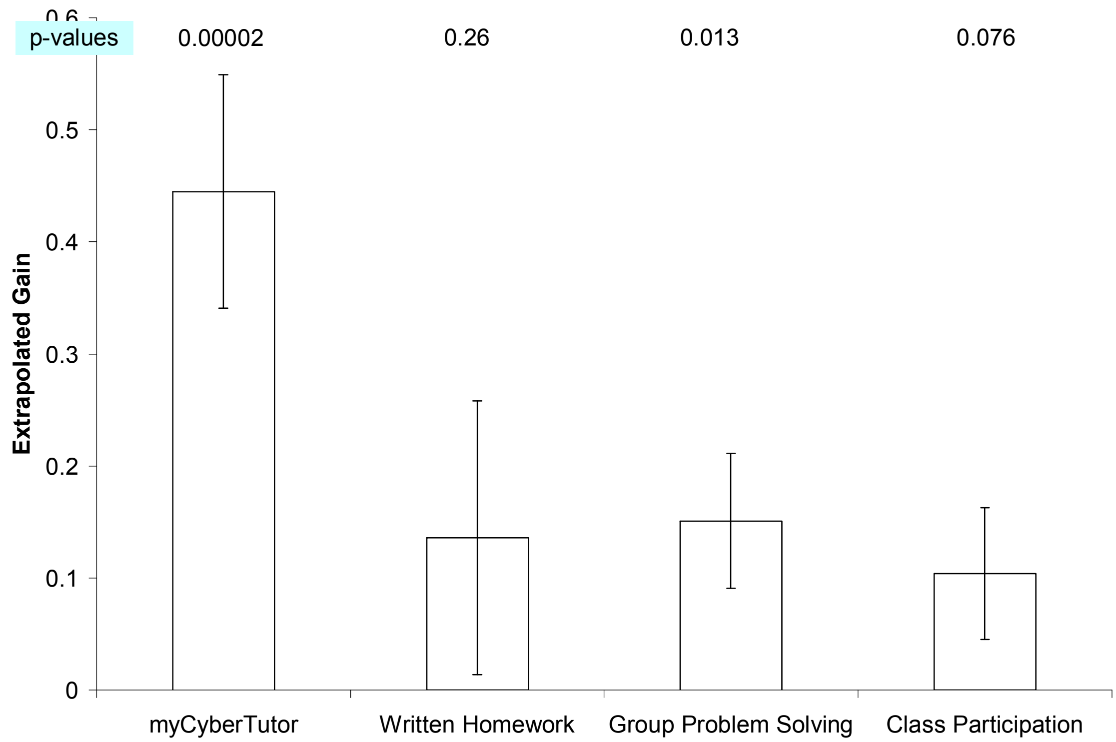

FIG. 2: Extrapolated gain on the MIT final exam vs various course elements, weighted average of results for 2001 and 2002 (see Table IV).

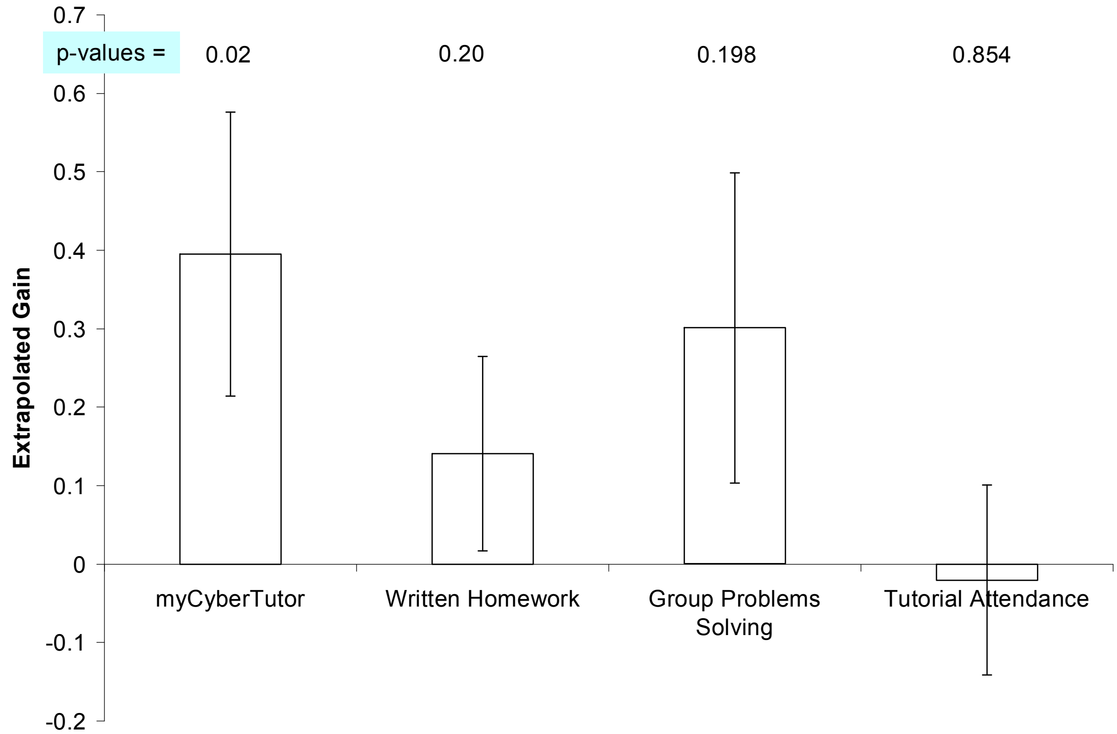

FIG. 3: Extrapolated gain on Force Concept Inventory vs various course elements – 2000.

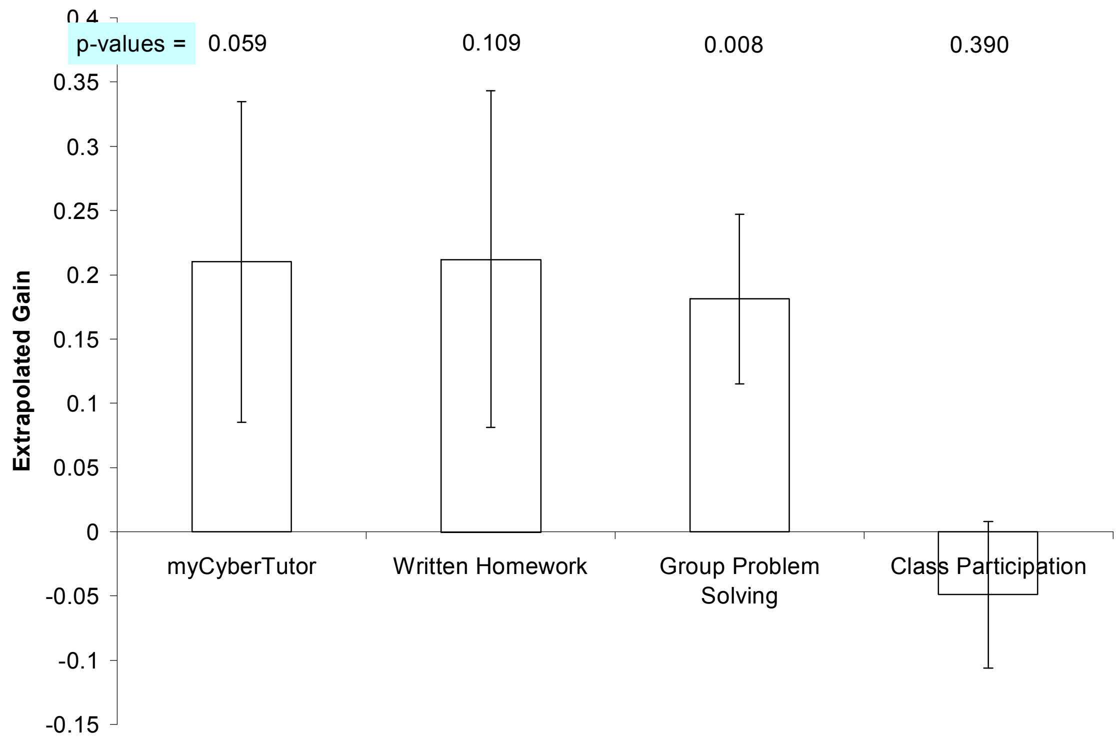

FIG. 4: Extrapolated gain on the Mechanics Baseline Test vs various course elements -2001.

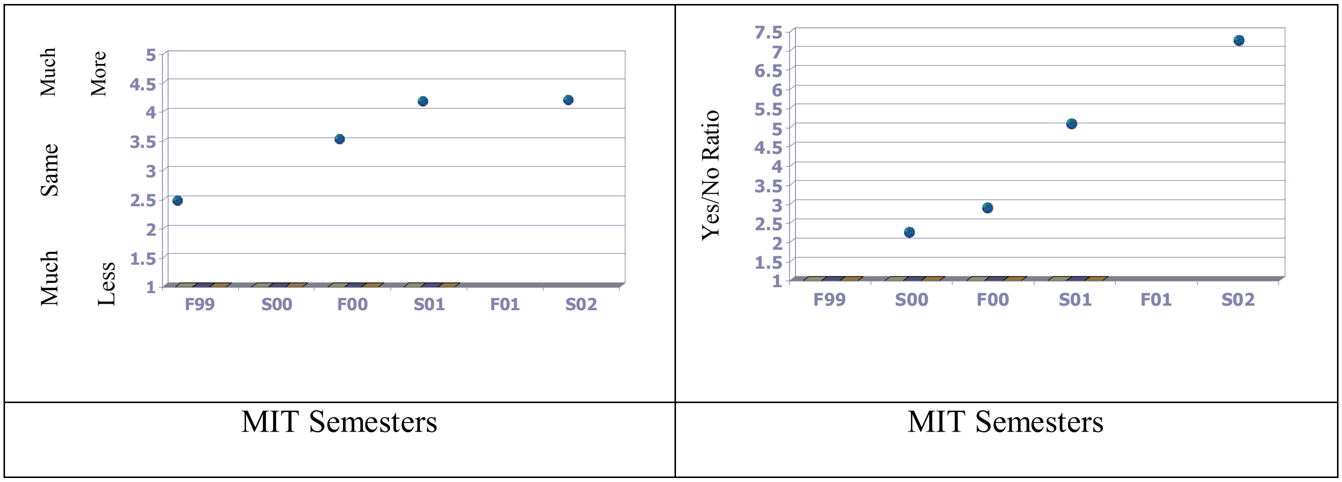

FIG. 5: Left panel: Average student response to "compare the amount you learn per unit time using myCyberTutor with time spent on written homework (including studying the solutions)." Right panel: Ratio of "yes" to "no" student responses to the question "Would you recommend myCyberTutor for use in 8.01 next year?" ("no opinion" responses were as few or fewer than "no" responses.)

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 9988732. We also acknowledge C. Ogilvie for comments on his study and discussions in the early phases of this paper. We acknowledge R. Warnakulasooriya for help with the manuscript and for comments on the paper.

[1] J. Van Aalst, "An introduction to physics education research," Can. J. Phys. **78**, 57-71 (2000).

[2] D.R. Sokoloff and R. K. Thornton, "Using interactive lecture demonstrations to create an active learning environment," Phys. Teach. **35**, 340-347 (1997).

[3] C. Crouch and E. Mazur, "Peer instruction: Ten years of experience and results," Am. J. Phys. **69**, 970-977 (2001).

[4] P. Heller and M. Hollabaugh, "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups," Am. J. Phys. **60**, 637-644 (1992).

[5] P. Heller, R. Keith, and S. Anderson, "Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving," Am. J. Phys. **60**, 637-644 (1992).

[6] MyCyberTutor was made by Effective Educational Technologies. The problems were written by David Pritchard using the software developed by Alex Pritchard. David Pritchard is now a consultant to Pearson which markets an improved version of this product - see www.MasteringPhysics.com.

[7] R. Hake, "Interactive engagement versus traditional methods: A six thousand student survey of mechanics test data for introductory physics courses," Am. J. Phys. **60**, 64-74 (1998).

[8] D. Hestenes, M. Wells, and G. Swackhammer, "Force concept inventory," Phys. Teach. **30**, 141-158 (1992).

[9] D. Hestenes and M. Wells, "A mechanics baseline test," Phys. Teach. **30**, 159-166 (1992).

[10] J. Saul, *Beyond problem solving, evaluating introductory physics courses through the hidden curriculum.* Ph.D. Dissertation, University of Maryland (1998)

[11] L. McDermott and P. Shaffer, *Tutorials in Introductory Physics*, **1**st edition (Prentice Hall, New Jersey, 2002).

[12] P. Laws, *Workshop Physics Activity Guide* (Willey, New York, 1996)

[13] C. Ogilvie, Effectiveness of different course components in driving gains in conceptual understanding. Internal report, Department of Physics, MIT (2001). (URL: http://relate.mit.edu (2000-2004).)

[14] H. Cooper, *Homework* (Longman, New York, 1989).

[15] J, Mestre, R. Dufrense, D. Hart, and K. Rath, “The effect of web-based homework on test performance in large enrollment introductory physics courses,” J. Comp. Math. Sci. Teach. **21**, 229-251 (2002).

[16] M. Thoennessen and M. Harrison, “Computer-assisted assignments in a large physics class,” Computers. Educ. **27**, 141-147 (1996).

[17] S. Bonham, R. Beichner, and D. Deardorff, “Online homework: Does it make a difference?” Phys. Teach. **39**, 293-296 (2001).

[18] Jacob Cohen, *Statistical power analysis for the behavioral sciences*, **2nd** edition (Lawrence Erlbaum Associates, New Jersey, 1988), p.20-25.

[19] David J. Palazzo, *Patterns and consequences of copying electronic homework* (Masters Thesis in Physics, Massachusetts Institute of Technology, 2006)

20. Elsa- Sofia Morote and David E. Pritchard **What Course Elements Correlate with Improvement on Tests in Introductory Newtonian Mechanics?**

Educational Resources Information Center (ERIC), ED463979