A comparison of student performance using web and paper-based homework in college-level physics.

Scott W. Bonham
Department of Physics, Western Kentucky University
Duane L. Deardorff
Department of Physics, University of North Carolina
Robert J. Beichner
Department of Physics, North Carolina State University

Abstract: Web-based homework is becoming widespread practice in physics and other technical courses, primarily for logistical and economic reasons. Although people argue whether this is a positive or negative development for student learning, there is limited research assessing the pedagogical effect of changing the medium from written, hand-graded homework to web-submitted, computer-graded work. Towards this end, we assessed student performance with web-homework and with traditional paper homework in a realistic classroom setting for large-enrollment service courses. One of two nearly identical sections of introductory physics received paper-based, hand graded homework while the other received the standard web-based homework, and we compared the two groups on conceptual and problem-solving performance measures. We conducted this experiment two times, once in a calculus-based course and once in an algebra-based course. No significant differences in student performance were found that could be attributed to the homework method used. This work was supported by the Spencer Foundation and the National Science Foundation under grant 9714546.

Introduction

Web-based homework is a rapidly growing educational use of the Internet. At least a hundred thousand U.S. students currently submit their homework for computerized grading over the Web while attending real (non-virtual) classes, and the practice is also growing rapidly in math, chemistry and other sciences*. In addition to this are students enrolled in on-line courses and those who use on-line practice quizzes and the like. "Anytime, anywhere" computerized systems which instantly mark answers right or wrong and then allow errors to be corrected are replacing traditional paper homework handed in once a week, graded by the instructor or an assistant, and returned with marks and comments days to weeks latter. Clearly a transition such as this could have significant ramifications, particularly in courses like physics, where homework is considered to be a crucial component for learning. Web-based homework also opens up new possibilities for types of exercises that are impossible in paper format, such as incorporating animations and simulations (Christian & Titus, 1998). However, there is scant research on this use of technology, even though it is needed to both guide adoption of web-based homework and to distinguish the pedagogical effects of the computerized medium from innovative web-based content. The latter is an important research question, as Weller (1996) points out the importance of separating out pedagogical gains because of the technology itself from that due to improved content or more time on task.

There are a number of motivations for employing web-based homework in large university courses. Benefits include providing students with immediate feedback, using the latest

* WebAssign currently serves about 40,000 students (John Risley, personal communication). Another system, Homework Service, boasts on their website (http://hw.ph.utexas.edu/hw.html) that they process 150,000 answers a week; at 5-10 questions per student, this works out to 20,000-50,000 students. In addition to this are schools using CAPA, OWL, WebCT, and other systems.

technology for instruction, reducing the grading burden on faculty and assistants, being able to grade all student work, and reducing costs by using computers instead of graduate assistants. The subscription to a full-service web homework service for a large service course can cost as little as one-fifth of the pay and benefits for a human grader (and direct institutional costs may be much less)*. However, there are potential drawbacks of using web-based homework, including a lack of detailed feedback to students, the danger of multiple submissions encouraging lazy habits, and further impersonalization of the course by replacing a human grader with a computer. Since web-based homework systems of the kind studied in this paper have been developed primarily by technologists and/or faculty in a particular subject matter, it is perhaps not too surprising that limited research exists. By and large web-based homework systems deliver fairly standard materials and exercises, changing the delivery medium but leaving the pedagogical content much the same, since the same exercises or types of exercises are delivered through the new medium. This contrasts with many excellent examples of computer assisted instruction, which use the computer as a medium for delivering pedagogically innovative materials. The focus of this work is the impact on learning coming from a change in the medium—what difference does it make to use computers and the Internet as opposed to paper and human graders to deliver, collect and grade introductory physics homework? This is an important question, both in practical terms for those using or contemplating the use of web-based homework, and in the larger research agenda in distinguishing the effect of the medium from content and pedagogy.

This paper comparing web-based homework and paper-based homework in multiple courses at a large state university will focus on student performance and quantitative measures of student learning as a function of how they did their homework. For the purposes of this paper, 'web-based homework' consists of assignments delivered, collected and graded over the Internet through a web-based homework system and which forms the homework component of a standard course. A 'web-based homework system' is a web-based service which (1) can be accessed from any standard browser and Internet connection (2) password authenticates the user, (3) delivers assignments to students and receives their answers, (4) grades student work automatically and (5) keeps a permanent record of student scores which the instructor can access at a latter time. A few of the currently available systems that meet this broad definition of 'web homework' include WebAssign, CAPA, Homework Service, WebCT, Blackboard, OWL, and WWWAssign (Blackboard, 2001; CAPA, 1999; Hart, Woolf, Day, Botch, & Vining, 1999; Mallard, 1999; Martin, 1997; Moore, 1997; WebAssign, 1998; WebCT, 2000). 'Computer-based homework' is a more general term for any type of homework graded by a computer, including web-based homework. 'Paper-based homework' is the more traditional method of students working out their solutions on paper, turning these in for grading (perhaps superficially, perhaps in-depth), and, after a delay of a few days to a few weeks, receiving the papers back with written comments on them. We focused on courses where the instruction took place in real (non-virtual) classrooms and where the assignments consisted of standard exercises (i.e. the kind found at the end of the textbook chapter). Note that the subject of this paper is more limited than much of the work in computer-aided instruction (CAI). Here we are dealing with the situation in which instruction is provided by regular classes and/or textbooks, and the computer is simply used for further practice

* For example consider a graduate student receiving a stipend of $14,500 a year (not an unreasonable figure in science and engineering departments). Five months of this salary is $14,500×5/12 = $6,050, not including tuition waivers and other benefits. If one were to grade all homework in a 100-student course, a grading assistant working half-time (20 hours/week) is about what would be needed. One of the major homework services in physics, WebAssign, offers a hosted service with a large databank of problems already entered and currently has a pricing scheme with a base price per instructor-semester of $75-$125, plus $8.50/student charge. This means that a 100-student course would be less than a thousand dollars per semester. Furthermore, the institution may have the students directly pay the per-student charges, meaning that the institution directly pays only a small fraction of that.

of already-learned material. Good CAI is a pedagogical strategy that utilizes a technological medium, which can be delivered by web homework systems. However, this work is a look at actual practice and a comparison of the effect of the medium using the same pedagogical content and strategy, differing only in aspects intrinsic to the medium—i.e. quality and timeliness of feedback.

In a typical web-based homework system, students log on using a password through the Internet to a central web server, select one or more assignments, and receive those exercises. A screen shot of such an assignment may be seen in Figure 1. In many cases the numerical exercises are randomized, so each student assignment has a unique set of numbers. Depending on the system and the individual, the students could work through the exercises while seated at the computer or they may print the exercises in order to work them out on paper elsewhere. After determining the answers, the student will then submit their solution, which is most commonly a numerical result or one option from a multiple choice list, but could also consist of selecting multiple options in a list, entering a symbolic answer (e.g. “d+v*t+0.5a*t^2”), typing in a word or a short essay, or uploading a file. In most cases, the computer immediately evaluates the answers, gives the student some level of feedback, and—depending on how the instructor has set options—allows reworking and resubmission of the assignment. The instructor is able to handle administrative details, create assignments and questions, and review or download student scores and responses. Some systems have additional features such as chat rooms, instructor notes, calendars and so forth. A detailed overview of web-based homework systems may be found in Titus, Martin, & Beichner (1998).

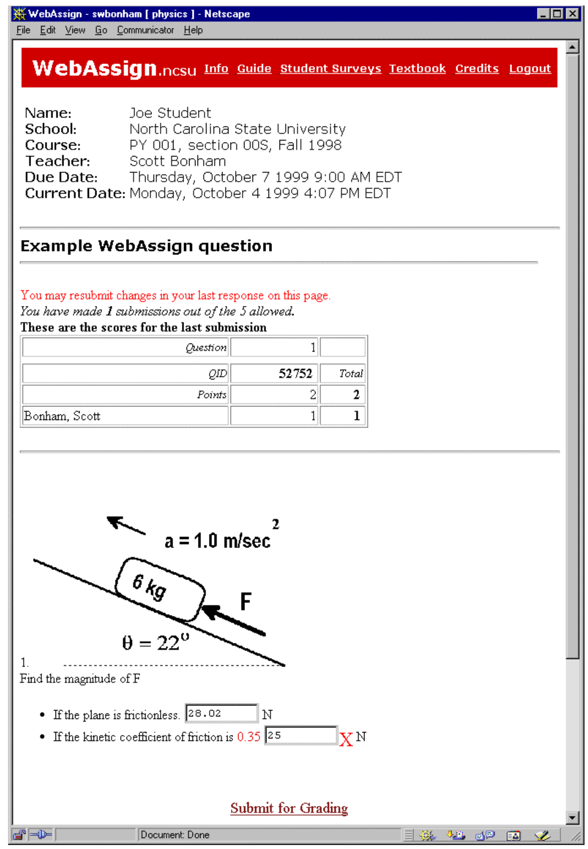

Figure 1: Screen shot of student page in WebAssign. Some numbers, such as the coefficient of friction in this example, are randomized so students get different numbers. This is signaled by the randomized numbers being red (in printed form this appears gray). Assignments usually consist of more than one question.

The roots of computerized homework systems in physics go back at least to the PLATO system (Sherwood, 1971) utilizing then-current technology, from mainframes with terminals or punch cards (Connell, 1994; Taylor & Deever, 1976) to personal computers and local networks (Abbott, 1994; Milkent & Roth, 1989) to the Internet and World-Wide Web (Kashy et al., 1993; Moore, 1997; Raineri, Mehrtens, & Hübler, 1997; WebAssign, 1998). This development has paralleled instructional technology advances in math, chemistry and engineering (Barker, 1997; Connolly, 1972; Graham & Trick, 1997; Hart et. al., 1999; Kohne, 1996; Maron, Ralston, Howe, & Townsley, 1999; Morris, 1982; Porter & Riley, 1996; Spain, 1996; Steinley, 1986; Woolf, Hart, Day, Botch, & Vining, 2000; Yaney, 1971). Studies almost invariably report very positive reactions to computerized homework (Connell, 1994; Jones & Kane, 1994; Kashy et al., 1993; Ory, 1997; Taylor & Deever, 1976); students like the immediate feedback and being able to resubmit assignments, while their instructors like not having to manually grade student work. However, research on the effectiveness of computerized collection of student work in physics and other subjects is more limited and often inconclusive. One study using large introductory physics lecture sections (Jones, Kane, Sherwood, & Avner, 1983) found students using PLATO to perform better than those who didn't, but there were also differences in instructors, supplemental instruction, and whether or not homework was collected. A study looking at two physical science classes taught by the same instructor (Milkent & Roth, 1989) found little or no difference between the section completing homework with a BASIC program developed by the instructor and those doing the same work on paper. In introductory statistics a difference was found between a section working on standard textbook problems and a section which used a drill program (Porter & Riley, 1996), but the program was written by the instructor, was used in mastery mode and gave fairly extensive feedback. In the limited research where there was strict replacement of traditional homework with computerized grading, the effect was not large enough to be significant given the limited statistics, and in the cases where a difference was found it could potentially be attributed to differences in instruction, content, and/or time-on-task.

The current literature does not really answer questions being raised about computerized homework, web-based or otherwise. Homework is quite important in technical courses such as introductory physics, where problem solving is a major focus and homework is the main venue for practicing this. Many students struggle to develop problem-solving skills in physics (Maloney, 1994), although directed instruction and feedback has been shown to be effective (Heller & Reif, 1984; Heller & Hollabaugh, 1992). In this paper we will look at the following questions:

- Does one medium (web or paper) lead to better conceptual understanding?
- Does one medium help students develop better problem-solving skills?
- Does one medium lead to differences in other aspects of the course, such as laboratories and seeking out help with exercises?

Method

In order to answer these questions, we carried out side-by-side comparisons of student performance in multi-section, large enrollment introductory physics courses. This investigation was carried out at North Carolina State University (NCSU), a land-grant institution with a large population of engineering students. The research method was a quasi-experimental design, in which an instructor assigned to teach two lecture sections of the same course agreed to cooperate with the investigators. One of the paired sections received their homework via WebAssign where it was graded by a computer. The other section wrote out solutions to their homework exercises on paper. These exercises were turned in and graded by a full-time (15-20 hours a week) graduate student grader. This is a far more thorough grading effort than often provided in large introductory physics classes; before the development of the WebAssign homework system, NCSU instructors were provided roughly 5 hours/week of student grading help. This would have been enough to grade one or two problems in an assignment, but not all of them. The paired sections met in the same lecture hall in adjacent time slots. Students registered for the two different sections through the standard course registration system and were unaware of the homework method until it was announced the first day of class. During the first few weeks of the semester they were able to switch into other sections if they wished. (There were no reports of anyone switching sections solely because of the homework arrangement.) Students had a two-hour laboratory every other week, which was taught by teaching assistants (TAs) who reported to the laboratory coordinator and not directly to the course instructors. Laboratory sections were not coordinated with lecture sections, so a laboratory section would have students from different lecture sections, and vice-versa. The on-campus Physics Tutorial Center (PTC) offered drop-in tutorial assistance by its staff of graduate and upper level undergraduate physics students, as well as providing a library of instructional software and videos. The university also provided a peer-instruction program known as Supplemental Instruction sessions, in which an advanced undergraduate student would be assigned to a particular course, would take notes in lectures, and then host regular sessions outside of class where students could get help. We carried out this experiment two times: once in the first-semester calculus-based physics course and then in the first-semester algebra-based physics course. Because the two experiments were very similar in methods and results, we will present them in parallel in their respective sections.

Experiment 1

The first investigation took place in the spring of 1999 in the first-semester calculus-based physics, an on-sequence semester for introductory engineering students. This course typically has 400-900 students enrolled in any given semester. There are multiple lecture sections of 80-110 students, taught by different instructors. During the semester of the study there were five additional sections taught by other instructors, for a total of seven. The population is primarily students in the engineering sequence, and the course covers the topics of kinematics and dynamics, rotational motions, oscillations, and waves. A course coordinator sets the syllabus, makes the default homework assignments (which few instructors change) and writes the common exams for all the students. The textbook for the class was Fundamentals of Physics, 5th ed. (Halliday, Resnick, & Walker, 1997). There were four common tests during the semester, constructed by the course coordinator. These consisted of 15 multiple-choice questions and a worked-out problem broken into several parts, accounting for 75% and 25% of the total points, respectfully. Homework and laboratories each counted 10% of the final course grade, the four mid-term exams combined for 56% and the cumulative multiple- choice final exam for 24%. Nearly all of the assigned homework problems were from the textbook. The web section received the standard WebAssign homework assignments made by the course coordinator that were given to all the other sections. The department had previously switched to entirely web-based homework for this course, so in this case the paper section is the treatment group, and the web section the control. The professor, a very experienced instructor who makes teaching his main focus, spent the majority of class time working problems similar to homework exercises and material on the exams. Many days there would be a time during class where students worked for 5-10 minutes in self-selected groups on one or more exercises. The professor taught the section using WebAssign at 12:25 PM Monday, Wednesday and Friday, and the section using paper-based homework at 1:30 PM on the same days. The WebAssign section generally had three weekly assignments due at 11:30 PM on Mondays, Wednesdays and Fridays, typically consisting of two or three questions from the text. The paper section submitted paper homework once a week, usually at the end of class on Friday.

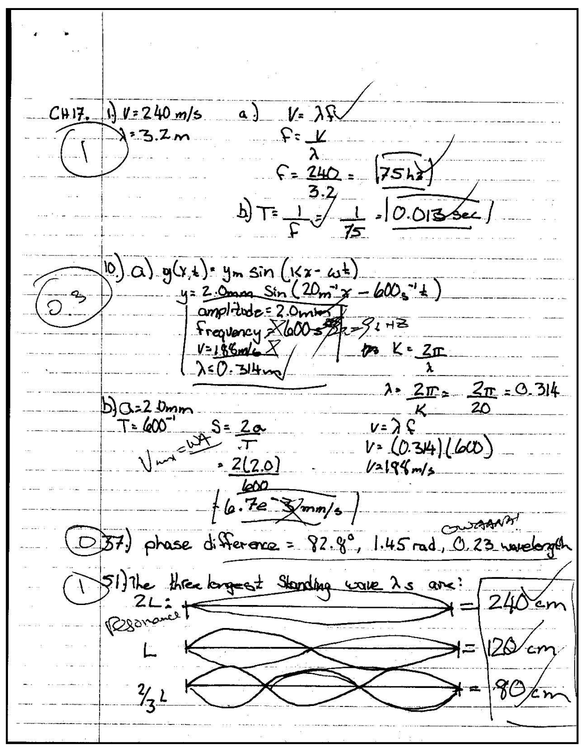

Figure 2: Example of grading from calculus-based course. The grader checked different parts but did not write a lot of comments. (English was not his native language.)

These students were asked to write solutions that included (a) identifying the information given in the problem (b) a drawing (c) a layout of the solution (the formulas) (d) the solution, complete with units and significant figures, and (e) a check for reasonableness. All problems were graded by a graduate student who spent up to 20 hours grading each week, including checking individual parts of problems. An example of grading may be found in Figure 2. Homework was returned through drop-off boxes at the PTC, which is located adjacent to the room where the students conducted their laboratory exercises. Most of the exercises the two groups worked were the same (or in a few cases, very similar) problems from the text and had the numerical answers in the back of the book. The web section also received via WebAssign an old exam as a practice test before each mid-term test; this was not distributed to the paper group, but old exams were readily available on the course website and in test packs from the bookstore. The paper group also turned in answers to a few conceptual exercises on each assignment, which the web students could submit on paper for a small amount of extra credit.

Experiment 2

In order to verify whether the results observed with the calculus -based course could be a result of the small difference in which exercises were assigned, the use of a non-native English speaking grader, or the particular population, we repeated the experiment in an algebra-based physics class in the fall of 1999. The first-semester algebra-based course has approximately 200-400 students per semester, taught in sections of 60-90 students by multiple instructors. It covers the same topics as the calculus-based course and is predominantly populated by biology and allied health science students. Unlike the calculus-based course, there was no common syllabus, homework or exams for all sections of the algebra-based course. In addition to the paired sections that participated in this study, there were three other sections of the course taught by other instructors, which will not be discussed further. As in the first experiment, students registered for the course with no knowledge of the homework policy, which was announced the first day of class. Students were able to register for a different open section if they chose to do so. The web and paper sections met on Tuesdays and Thursdays at 11:20 AM and 1:05 PM respectively. Most weeks there was a quiz given in class with one or two problems very similar to a homework exercise, where written-out solutions were required. These quizzes were open book and open notes. Students were able to "recover" up to half the points lost on the quizzes by going to the PTC to rework the quiz and complete several additional problems. There were three multiple-choice exams during the semester written by the instructor. Quizzes counted for 40% of students' grade, laboratories for 10%, homework for 9%, tests for 40%, and 1% for simply logging into the instructor's fairly extensive website. The main focus of the instructor's activity during the lectures was working through the assigned homework problems, frequently leaving the final

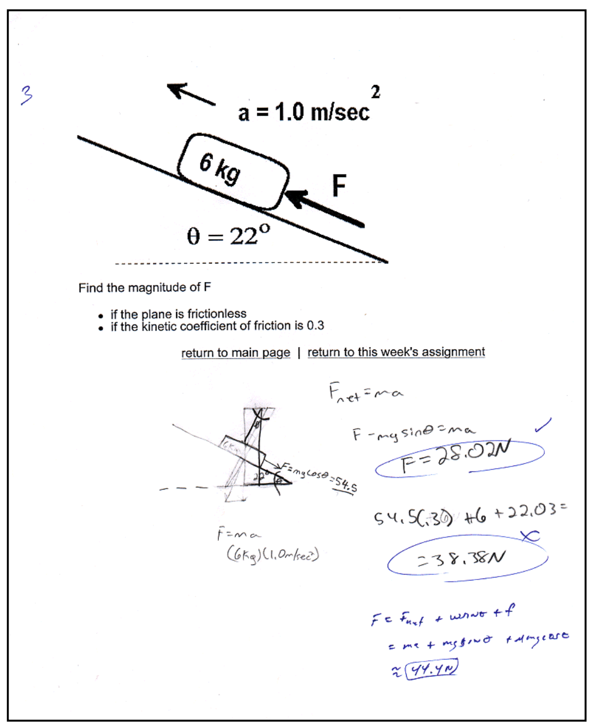

Figure 3: Example of grading from algebra-based course. This grader gave more extensive feedback. In this case, it appears that the student made a calculation error.

numerical calculations for the students. Although the department had designated College Physics (Serway & Faughn, 1999) as the official text, the instructor chose not to tie the course closely to any particular textbook, and so wrote all of the homework exercises in the style of typical end-of-the-chapter problems. The assignments consisted of ten to twelve exercises each week, and usually several of the problems were multi-step. Most of the problems were numerical, but some multiple choice and essay questions were also used. These were made available to all students via the course website, delivered via WebAssign to the web section, and handed out in printed homework packets—one problem per page—during class in the paper section. Both sections received exactly the same weekly assignments, which were due once a week at nearly the same time. For a number of years this instructor had not collected homework in either electronic or paper form due to lack of grading help. Difficulties with the web system at the beginning of the course caused the due time for the assignment to be pushed back several times in the first few weeks, finally being fixed at 9 AM Thursday morning for both sections. Students in the paper section were required to show work on the paper assignments, which were graded by hand and returned in class, 1-2 weeks later. The TA for this section was an American physics graduate student who did a thorough job of grading, including giving positive as well as negative feedback. An example of grading may be seen in . At the beginning of this particular semester the WebAssign service experienced technical difficulties which made the system sluggish and unresponsive for several weeks. There were also some errors in coding the answers to the instructor-written problems due to the short lead-time available, so the first students to work the problems sometimes found exercises marked incorrectly before the problem was discovered and fixed. As a result, the instructor and many of the web homework students developed a negative attitude towards the system over the course of the semester. While not done deliberately, the circumstances of the two investigations span a range of implementation scenarios, with the calculus course possibly biased more favorably toward the web-based section while the algebra course ran much smother for the paper-based section.

Results

Experiment 1

We collected data on test performance, homework scores, laboratory scores, a pre/post conceptual test, utilization of the PTC, and in-class survey and interviews. Scores and responses on the multiple-choice part of the test were obtained from the course coordinator and worked-out problems were photocopied before they were returned to students. The Force and Motion Concept Exam, (FMCE, Thornton & Sokoloff, 1998) was administered to all students in the course—including those in this study—in their laboratory sections at the beginning and end of the semester. Students received extra credit for participating but it was not required; most students participated at the beginning and about half of all students participated at the end. The multiple-choice FMCE probes conceptual understanding of physics, particularly the degree to which students hold to Newtonian as opposed to non-Newtonian beliefs. The values reported here are raw (percent correct) scores on this test. The university registrar provided grade-point average (GPA) and scores on the Scholastic Aptitude Test math section (SATM). A subset of the solutions to the written part of the exam were carefully analyzed for differences following the markup procedure of Stewart (in press); this is a quantitative method of describing the kind of solution and quantifying different aspects of the solution such as the number of equations, number of symbols (variables), numbers of figures, labels on the figures, numbers with units, etc. The use of the PTC, the survey, and interviews will be discussed in the second paper.

Data from the different sources was compiled together, and students who did not complete the course were removed; completing the course was defined as those who received a final grade (did not withdraw) and took at least one exam. There were a total of 117 students (35 women) in the web section and 113 students (20 women) in the paper section. Table 1 summarizes the comparison between the web and paper sections, using two-tailed t-tests. Because not all background information was available for all students, the *N* is smaller for some items. GPA, SATM and FMCE pretest give background information on the students and allow us to judge how well matched the paired sections were. From these measures, we can see that the section doing web-based homework entered at a slightly higher academic level, but in no category was the difference significant at the $p < 0.05$ (95% confidence) level, so the sections were relatively well matched.

Table 1
Comparison of background and performance measures in the calculus-based physics course.

Measure	Web section			Paper section			t-test	
	N	Mean	SD	N	Mean	SD	score	p
GPA (A=4.0)	110	3.11	0.61	108	3.01	0.68	1.16	0.25
SAT math score	111	639	81	109	632	69	0.68	0.50
FMCE pretest (%)	98	26.5	15.8	95	26.1	16.2	0.02	0.99
Homework average[a]	117	87.9	22.7	112	72.7	32.5	4.13	< 0.0001
Test Average	117	75.4	13.1	112	73.3	13.9	1.18	0.24
# MC questions correct	105	11.7	2.0	105	11.2	2.3	1.53	0.13
Written question points[b]	105	20.2	3.7	105	18.9	4.2	2.21	0.03
FMCE gain (%)	60	18.9	24.3	38	20.1	28.6	0.06	0.95
Lab average	117	84.9	17.2	112	84.3	14.2	0.23	0.78

Mean, standard deviation and results of two-tailed t-test assuming unequal variances. GPA, SAT and FMCE data not available for all students. Conflict exam data is included in Test Average but not MC questions and Written questions.
[a] Score on final submission for web section, only submission for paper section.
[b] The two sections had different graders for the written part of the exam.

Student performance data is compared in the remaining portion of Table 1. Homework average is the actual homework score obtained divided by the total possible points. The web students were allowed to resubmit problems without limit until the deadline, and their score represents their final (not necessarily best) submission. The paper students were only able to submit once, but their score includes partial credit. The web students in the calculus section also had three short assignments a week, while the paper group had a single, longer assignment. It is therefore not surprising that in the calculus course the web students had a higher homework score. The calculus course had four mid-term tests and a cumulative final. The tests had 15 multiple-choice questions and a multi-part written question that was graded by graduate students (different graders for the two course sections). The final was entirely multiple-choice. The average reported here is the average of the four tests and the final, with the final having twice the weight of a mid-term test. "MC questions" are the average number of multiple-choice questions a student got right on the regular mid-term exams. (Make-up exams for students who could not attend the common exam time were scored differently and were not available.) "Written questions" is the average score on the written section of the regular mid-term exams. There is a statistically significant difference between the treatment and control groups on the written questions, but this measure can not tell us how much it is due to differences in the sections versus differences in scoring between the two graders. The astute reader may notice that the t-test statistic for both the

MC questions and written questions is higher than the statistic for the test average. The difference between the two sections on the final, which is not included in the first two items, was very small. In the calculus course the FMCE was given again at the end of the semester. Participation was voluntary (extra credit given) and about half of the students showed up. For those who participated in both pre- and post-testing, we can calculate a gain ratio as the fraction of possible gain actually achieved, computed as (Hake, 1998)

$$g = \frac{posttest - pretest}{100\% - pretest}.$$

There is no significant difference between the web and paper students for the FMCE gain in the calculus course. The laboratory scores are the scores reported by the laboratory teaching assistants (TAs) on student lab reports; although there may be variation from TA to TA, students were mixed in the laboratory sections, so possible grading difference would tend to average out, and no significant difference between the treatment and control groups is apparent. In summary, the web section proved to have slightly (but not significantly) better students than the paper section, and on some of the measures of class performance, they out-performed their paper colleagues at a significant level, but it is not clear just how much of that is due to the homework type.

Table 2
Summary of Hierarchical Regression Analysis for variables predicting score on written part of exam in the calculus-based course.

	Measure	B	SE B	β	p
Step 1					
	Intercept	-0.45	2.08	-0.02	0.83
	GPA	2.57	0.39	0.41	$< 10^{-9}$
	SAT math score	0.016	0.003	0.53	$< 10^{-6}$
	FMCE pretest	0.063	0.015	0.25	$< 10^{-4}$
	Paper class	-0.16	0.44	-0.01	0.71
Step 2					
	Intercept	-1.26	2.15	-0.05	0.56
	GPA	1.73	0.42	0.28	0.00005
	SAT math score	0.016	0.003	0.51	$< 10^{-5}$
	FMCE pretest	0.069	0.015	0.28	$< 10^{-5}$
	Paper class	0.30	0.44	0.01	0.49
	Homework average	0.04	0.01	0.16	$< 10^{-4}$
	Male	0.11	0.52	0.00	0.83
	Minority[a]	-0.87	0.86	-0.03	0.31

N = 172. For step 1, $R^2 = 0.47$ ($R^2_{adj} = 0.46$). For Step 2, $R^2 = 0.53$ ($R^2_{adj} = 0.51$)
[a] A member of an under-represented minority: African American, Hispanic, or Native American.

In order to explore further the relationship between test performance and background factors, a regression analysis was carried out on written question scores, which were worth 25 points. A hierarchical regression was carried out, first using the background factors of GPA, SAT math scores, scores on the FMCE pretest, and whether the student had paper homework (the treatment). A summary of this analysis may be found in Table 2. GPA, SAT and FMCE were very strong predictors of performance on the written questions, accounting for nearly half the variance and with p values of less than 0.0001, while homework method was insignificant. In a second step, homework scores, gender and minority status were added in. Of these additional factors, only homework average made a significant contribution, but even so its contribution to the model, as measured by the standardized coefficient β, is much less than that of GPA, SAT or FMCE. A similar result for the average number of multiple-choice questions correct on a test is shown in Table 3. This shows us that the difference seen in the t-test results for the written questions is attributable to pre-existing differences between the two groups—e.g. math skills, previous knowledge of physics and general level of student—and not from the difference in how homework was done.

Table 3
Summary of hierarchical regression analysis for variables predicting score on multiple choice part of exam in the calculus-based course.

Measure	B	B SE	β	p
Intercept	-1.96	1.01	-0.13	0.05
GPA	1.28	0.21	0.34	$< 10^{-8}$
SAT math score	0.011	0.002	0.59	$< 10^{-10}$
FMCE Pretest	2.97	0.72	0.20	$< 10^{-4}$
Paper class	0.28	0.22	0.02	0.20
Homework average	0.017	0.005	0.12	< 0.0005

N = 172. $R^2 = 0.60$, $R^2_{adj} = 0.59$.

An additional question is if there were differences in skills related to problem solving that are not reflected in the exam scores. The students who worked their homework on paper received more practice in writing out problems on paper, since they had to show all work, so there might be a difference in how the two groups wrote out their solutions on exams. One of the concerns about computer-based homework is that it further reduces the incentive for students to write systematic solutions, explaining steps, working algebraically, keeping track of units, and so forth. Writing systematic solutions is good practice for students to learn, since it both helps to communicate clearly to others what was done and can help in preventing errors. Clearly labeling quantities—including using words—can help avoid later confusion about what is what. Mistakes can be avoided or more easily caught by working algebraically through the solution step by step instead of jumping steps or substituting numbers as soon as possible. Including units in all the calculations and doing a unit check at the end is also a valuable error-checking procedure. It might have been possible that the paper section developed better skills in these areas, which was not reflected in test scores.

In order to determine if there were any substantial differences between the paired sections, student solutions on a written solution were analyzed. All written solutions were photocopied after being graded and before being returned to the students. After looking at several different exercises, we decided to analyze in depth the final part (part E) of the written section of the second exam. This exercise was chosen because it is a multi-step exercise and involves core ideas of velocity, force and energy, so a well-organized solution can be very helpful in successfully solving it. Furthermore, this was the second exam so the students were already familiar with the exam format. In order to determine if there were quantitative differences between the two sections in verbal descriptions, numbers of equations, and use of variables, numbers and units, different solution components were counted for all the solutions (Stewart, in press). Students who took a make-up test took a different exam, so this analysis involved 82 solutions from the web students and 78 solutions from the paper section. The number of words, excluding words appearing as a subscript of a variable, were counted as well as the number of equation phrases. An "equation phrase" is a phrase in which a quantity is equal to another using either the word "is" or an equal sign, and at least one of the quantities is described verbally. Also counted were the number of equation signs, the total number of variables written, the number of numbers (excluding obvious arithmetic calculations in margins), the number of units, and number of answers boxed or circled. A complete, properly written solution would contain two boxed answers, but many solutions did not, either because the student was unable to complete the solution or did not follow instructions to box the answers. An example of a good solution to this exercise and the count on the various quantities is shown in Figure 5.

Table 4 summarizes the results from this procedure, giving the average of each group, t-test score and the p value. As can be seen from the table, no significant differences were observed at the $p < 0.05$ level and only boxing the answer was significant at the $p < 0.1$ level. This last item is perhaps the least significant of the problem solution elements, as it is strictly a mechanical communication element that makes it easier for the grader but does not contribute in any way to the solution process itself.

Table 4
Comparison of the numbers of elements in student solutions to last written exam question on second mid-term test.

Count of	Web section Mean	SD	Paper section Mean	SD	t-test score	p
Words	2.6	4.5	2.3	5.0	0.41	0.68
Eqn. phrases	0.4	0.7	0.3	0.5	1.09	0.28
Equation signs	7.4	3.7	7.5	3.3	-0.20	0.84
Variables	11.7	6.8	12.2	6.4	-0.53	0.60
Numbers	11.4	4.8	12.0	4.7	-0.08	0.42
Units	5.5	3.6	6.1	4.8	-0.88	0.38
Units/number	0.5	0.3	0.5	0.3	0.30	0.76
Answers boxed	1.1	0.9	1.3	0.8	-1.68	0.09

Solutions where nothing was written or student did not take the regular mid-term test were excluded, leaving a total of 82 solutions from the web section and 78 solutions from the paper section.

In summary, the only measurable differences in quantitative performance measures between the students working their homework on paper and those working their homework in the web-based computer system are directly attributable to differences in the populations themselves and not to the treatment. The only exception is homework score, but the difference in the number of times they were able to submit means that this measure is not really comparable between the groups. The substitution of human graded paper homework for computer graded web homework made no measurable difference in student learning.

Experiment 2

Most of the same data was collected in the algebra based physics course. Data was collected from the instructor on student performance on quizzes, exams, homework, and quiz make-ups. Selected background data on students was obtained from the university registrar. On two of the homework assignments students were also given a short survey about homework, and augmented data on use was obtained from the PTC. No course-wide FMCE testing was done that semester. Students not completing the course were removed from the analysis, leaving 64 students (37 women) in the web section and 56 students (35 women) in the paper section. Table 5 summarizes this data. As in the case of the calculus-based course, the web-based section had higher GPA and SAT math scores. We can not tell to what extent the tendency of better students to be in the web section is due to the type of homework, to being scheduled earlier in the day, or to random fluctuations.

Table 5
Comparison of background and performance measures in the algebra-based physics course.

Measure	Web section		Paper section		t-test	
	Mean	SD	Mean	SD	value	p
GPA[a]	3.19	0.79	2.96	0.49	1.52	0.13
SAT math score[a]	596	76	571	64	1.9	0.06
Homework	65.0	26.8	62.5	20.1	0.57	0.57
Test Average	84.2	17.5	77.3	14.3	2.35	0.02
Quiz Average	6.3	2.0	5.8	1.1	1.57	0.12
Quiz Reworks	2.3	2.2	2.9	2.4	-1.61	0.11
Lab[b]	81.6	13.9	81.7	15.0	-0.04	0.94

Mean, standard deviation and results of two-tailed t-test assuming unequal variances. Except as noted, N_{web}=64 and N_{paper}=56.
[a] Data not available for 6 students, so N_{web}=58.
[b] One student was exempt from lab, so N_{web}=63.

The algebra course had three (non-cumulative) tests during the semester, which were entirely multiple-choice, and weekly open-book quizzes consisting of written problems closely related to the homework exercises for that week. The same grader marked quizzes in both sections in the algebra course. "Quiz Reworks" refers to the number of times students utilized the policy that they could rework quizzes with additional exercises to earn half of the missed point on a particular quiz. The t-test comparisons show a significant difference ($p < 0.05$) in the test scores but not in the quiz scores. Also note that the difference between the two sections on the math section of the SAT, another mathematical, multiple-choice test is nearly at the $p<0.05$ level, so it is not obvious how much of the difference is due to the differences in homework and how much can be attributed to different ability levels of the students. It is also noteworthy that, unlike the experiment in the calculus-based physics course, the homework scores do not differ significantly in this case. A number of factors may have contributed to this lack of difference: both sections had a long homework assignment each week, the instructor substantially worked many of the homework problems in class before they were due, and web students experienced technical frustrations with the system.

As in the case of the calculus-based course, a linear regression analysis was performed on the test and quiz performance data from the algebra class. A hierarchical analysis was undertaken for test performance, first including GPA, SAT and type of homework, and in the second step homework average, gender and minority status were included. This is summarized in Table 6. Once again, GPA and SAT math scores were strong predictors of performance while the type of homework and minority status were insignificant. Table 7 summarizes a regression analysis on quiz scores. As seen in the first experiment, student ability as demonstrated by GPA and SAT math scores are strong predictors of test and quiz scores. Homework average makes a smaller contribution, and the method of submitting homework does not make a significant difference.

Table 6

Summary of Hierarchical Regression Analysis for variables predicting score on tests in the algebra-based course.

	Measure	B	SE B	β	p
Step 1					
	Intercept	-8.99	10.58	-0.09	0.40
	GPA	13.06	2.21	0.52	$< 10^{-7}$
	SAT math score	0.086	0.018	0.69	$< 10^{-5}$
	Paper	-1.55	2.32	-0.02	0.51
Step 2					
	Intercept	2.50	11.49	0.03	0.83
	GPA	12.81	2.51	0.51	$< 10^{-5}$
	SAT math score	0.068	0.020	0.54	< 0.001
	Paper	-1.39	2.28	-0.01	0.54
	Homework average	0.058	0.056	0.06	0.31
	Male	5.91	2.41	0.06	0.02
	Minority[a]	-4.11	4.28	-0.04	0.34

N = 110. For step 1, $R^2 = 0.48$ ($R^2_{adj} = 0.46$). For Step 2, $R^2 = 0.51$ ($R^2_{adj} = 0.49$)

[a] A member of an under-represented minority: African American, Hispanic, or Native American.

Table 7

Summary of Hierarchical Regression Analysis for variables predicting score on quizzes in the algebra-based course.

Variable	B	SE B	β	p
Intercept	-1.44	1.00	-0.14	0.15
GPA	0.94	0.24	0.38	< 0.0002
SAT math score	0.005	0.002	0.42	< 0.005
Paper	-0.08	0.22	-0.01	0.73
Homework average	0.025	0.005	0.25	$< 10^{-5}$

N = 110. $R^2 = 0.52$, $R^2_{adj} = 0.50$.

Discussion

We have carried out a detailed study comparing the use of paper and computer homework in two different introductory physics courses. The two quasi-experiments involved two different populations of students; one consisting of primarily engineering students of whom a majority was male, and the other largely allied health majors and other sciences, of which a majority were women. The experiments also involved two different instructors and two different graduate student graders. Performance on tests, quizzes, conceptual tests and written solutions were analyzed. It was found that the student background, as measured by GPA, SATM and FMCE pretesting were significant predictors of student performance on the different measures, but homework method was insignificant in both experiments. Even looking at elements of written solutions on exams, we found no significant differences at the $\alpha = 0.05$ level. Thus, we conclude that we have established the null hypothesis that, in the case of introductory university level physics with standard lecture sections using typical end-of-the-chapter problems, there is no significant difference in student course performance between web-based homework with computer grading and homework written out on paper and graded by hand.

It is perhaps not so surprising that the difference in homework method has such limited effect on student performance. First of all, the pedagogical differences between the two methods are not very large. For the most part, the same end-of-the-chapter type problems were used, so there was no real difference in pedagogical content. The differences between the two homework methods are completeness required and feedback. The paper students were required to work out the entire solution and show their work, while the web students only needed to submit the final numerical answer. The paper students received more detailed information, but after a time delay, while web students received immediate feedback on whether their answers were correct or not. The paper students could check their answers with those in the back of the book and rework their solutions, while the web students could correct answers marked wrong by the computer and then resubmit. Furthermore, study practices of many students may tend to further reduce these differences. Many web students usually printed out assignments, worked them out on paper—sometimes very thoroughly—and then returned to the computer to submit them. Thus, many of the web students actually worked out their solutions on paper just as those in the paper section, simply using the computer as the place to check their answers and get credit. On the other hand, many of the students in the paper section did not spend much time reviewing the returned homework, viewing it as not important or not very helpful, and so did not derive as much benefit as they might have from the written grading. Both of these student habits tended to further reduce the effect of the differences between the two homework methods. The instructor and lecture style probably has little effect on this phenomena, since it has been observed that in the lecture-style classroom that still dominates most college level physics, teaching style and experience of the instructor have limited impact on student learning (Hake, 1998; Halloun & Hestenes, 1985).

This result also raises deeper questions that are beyond the scope of this paper about what the tests are actually measuring and the pedagogical value of standard textbook exercises. As noted above, homework score has less predictive power (as measured by the standardized coefficient β) of performance on both multiple-choice and written questions than do GPA, SAT or the FMCE pretest. In the algebra course, homework does not make a significant contribution at all to the model for test score. Even on quizzes, where the problems were similar or occasionally even identical to homework problems and students could refer to their homework solutions if they had already worked them out, both SAT and GPA have larger β coefficients than homework. One possible explanation is that these types of exams evaluate as much general study and test-taking skills as they assess physics knowledge. Another explanation might be that traditional end-of-the-chapter homework exercises do not really help students learn that much (Heller & Hollabaugh, 1992; Sweller, 1988; Sweller, Mawer, & Ward, 1982). The good news is that web-based homework opens up new possibilities for the types of homework exercises that could be used in the future—perhaps providing a way to incorporate the principles of well designed computer-assisted instruction into mainstream practice. A third possible explanation is that good students with a rigorous physics background not do not put a lot of effort into the homework—and thus do not receive good homework scores—because they realized that they did not need to do the homework to do well on the tests. One interviewed student did just this, but it is difficult to say how widespread this practice is.

Web-based homework is a rapidly growing use of the Internet and is becoming a major component of instruction in physics and other subjects. Web delivery and grading of traditional textbook-type questions is equally effective as having students write them out for hand grading by the instructor or grading assistant for student performance on exams and the layout of the solutions. This was the case in both calculus-based and algebra-based physics and with different instructors, and is consistent with the limited research that has appeared on this subject. We conclude that the change in the medium itself, from paper solutions and hand grading to answers in html forms with computer grading, does not have significant pedagogical consequences. Replacement of hand-graded homework by computer work could improve student learning by freeing time and economic resources for more effective instructional methods, and it could be a medium that allows widespread use of quality computer-assisted instruction to be incorporated into mainstream courses.

We would like to thank the Spencer Foundation for the grant that made this work possible, and the National Science Foundation for the fellowship that supported the lead author. People who helped make this work possible include Phillip Alexander, Qunyi Chin, John Hubisz, Louis Hunt, Larry Martin, George Parker, Nikhil Phansalkar, Brian Pittman, Elizabeth Reig, and John Risley.

References

Abbott, H. (1994). Physics Homework Management: Using Hypercard to Generated Individualized Problem Sets. Computers in Physics, 8(2), 166-169.

Barker, D. S. (1997). CHARLIE: A Computer-Managed Homework, Assignment and Response, Learning and Instruction Environment. Paper presented at the Frontiers in Education Conference (Am. Soc. Engineering Ed.).

Blackboard. (2001). Blackboard. Available: http://www.blackboard.com/.

CAPA. (1999). CAPA: A Computer-Assisted Personalized Approach. Michigan State University. Available: http://capa4.lite.msu.edu/homepage/.

Christian, W., & Titus, A. (1998). Developing Web-based Curricula Using Java Physlets. Computers in Physics, 12(3), 227-232.

Connell, J. H. (1994). Personalized Homework Pioneer. American Journal of Physics, 62(7), 585.

Connolly, J. W. (1972). Automated homework grading for large general chemistry classes. Journal of Chemical Education, 49, 262.

Graham, C. R., & Trick, T. N. (1997). An Innovative Approach to Asynchronous Learning using Mallard: Application of Java Applets in a Freshman Course. Paper presented at the Frontiers in Education Conference (Am. Soc. Engineering Ed.).

Hake, R. (1998). Interactive-Engagement vs. Traditional Methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64-74.

Halliday, D., Resnick, R., & Walker, J. (1997). Fundamentals of Physics. (5 ed.). New York: John Wiley & Sons.

Halloun, I. A., & Hestenes, D. (1985). The Initial Knowledge State of College Physics Students. Am. J. Phys., 53(11), 1043-1055.

Hart, D., Woolf, B., Day, R., Botch, B., & Vining, W. (1999). OWL: an integrated web-based learning environment. Paper presented at the International Conference on Math/Science Education & Technology (M/SET 99), San Antonio, TX.

Heller, J. I., & Reif, F. (1984). Prescribing effective human problem solving processes: Problem description in physics. Cognition and Instruction, 1, 177-216.

Heller, P., & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. American Journal of Physics, 60(7), 637-644.

Jones, L. M., Kane, D., Sherwood, B. A., & Avner, R. A. (1983). A final-exam comparison involving computer-based instruction. American Journal of Physics, 51(6), 533-538.

Jones, L. M., & Kane, D. J. (1994). Student evaluation of computer-based instruction in a large university mechanics course. American Journal of Physics, 62(9), 832-835.

Kashy, E., Sherrill, B. M., Tsai, Y., Thaler, D., Weinshank, D., Engelmann, M., & Morrissey, D. J. (1993). CAPA - An integrated computer-assisted personalized assignment system. American Journal of Physics, 61(12), 1124-1130.

Kohne, G. S. (1996). An Autograding (Student) Problem Management System for the Computer. Paper presented at the American Society of Engineering Educators Annual Conference.

Mallard. (1999). Mallard. University of Illinois. Available: http://www.cen.uiuc.edu/Mallard/.

Maloney, D. P. (1994). Research on Problem Solving: Physics. In D. L. Gabel (Ed.), Handbook of Research on Science Teaching and Learning (pp. 598). New York: MacMillan Publishing Company.

Maron, M. J., Ralston, P. A., Howe, T. E., & Townsley, M. B. (1999). Web-Based Grading of Symbolic Answers. Paper presented at the Frontiers in Education Conference (Am. Soc. Engineering Ed.).

Martin, L. (1997). WWWAssign. Available: http://www.admin.northpark.edu/lmartin/WWWAssign/.

Milkent, M. M., & Roth, W.-M. (1989). Enhancing student achievement through computer-generated homework. Journal of Research in Science Teaching, 26(7), 567-573.

Moore, C. F. (1997). Homework Service, [HTML]. Available: http://hw.ph.utexas.edu/overview.html.

Morris, J. (1982). Individualized, computer-managed homework assignment system (CS). Journal of Chemical Education, **59**, 603.

Ory, J. C. (1997). Student Use of and Attitudes about On-Campus ALN. Paper presented at the Frontiers in Education Conference (Am. Soc. Engineering Ed.).

Porter, T. S., & Riley, T. M. (1996). The Effectiveness of Computer Exercises in Introductory Statistics. Journal of Economic Education, 27(4), 291-299.

Raineri, D. M., Mehrtens, B. G., & Hübler, A. W. (1997). CyberProf - An Intelligent Human-Computer Interface for Interactive Instruction on the World Wide Web. JANL-Journal of Asynchronous Learning Networks, 1(2), 20-36.

Serway, R. A., & Faughn, J. S. (1999). College Physics. (5 ed.). Fort Worth: Saunders College Publishing.

Sherwood, B. A. (1971). Free-Body Diagrams (a PLATO Lesson). American Journal of Physics, 39, 1199-1202.

Spain, J. D. (1996). Electronic Homework: Computer-Interactive Problem Sets for General Chemistry. Journal of Chemical Education, 73, 222.

Steinley, C. D. (1986). Grading Assignments Is Enjoyable (So Long As It's Done For You!). Paper presented at the Frontiers in Education Conference (Am. Soc. Engineering Ed.).

Stewart, G. B. (in press). Using Quantitative Media Measurements to Manage, Improve, and Transport Introductory Science Education. Journal of Science Education and Technology.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. Cognitive Science, 12, 257-285.

Sweller, J., Mawer, R., & Ward, M. (1982). Consequences of history: Cued and means-end strategies in problem solving. American Journal of Psychology, 95(3), 455-483.

Taylor, J. A., & Deever, D. L. (1976). Constructed-response, computer-graded homework. American Journal of Physics, 44(6), 598-599.

Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. American Journal of Physics, 66(4), 338-351.

Titus, A., Martin, L., & Beichner, R. (1998). Web-based Testing in Physics Education: Methods and Opportunities. Computers in Physics, 12(2), 117-123.

WebAssign. (1998). WebAssign. North Carolina State University. Available: http://webassign.net/info/.

WebCT. (2000). WebCT. WebCT. Available: http://www.webct.com/.

Weller, H. G. (1996). Assessing the Impact of Computer based Learning in Science. Journal of Research on Computing in Education, 28(4), 461-485.

Woolf, B. P., Hart, D. M., Day, R., Botch, B., & Vining, W. (2000). Improving Instruction and Reducing Costs With a Web-based Learning Environment. Paper presented at the International Conference on Mathematics/Science Education & Technology (M/SET 2000), San Diego, CA.

Yaney, N. D. (1971). Computer system for individual homework. Keycard assembly, grading, and grade summation. Journal of Chemical Education, 48, 276.