The Effect of Web-Based Homework on Test Performance in Large Enrollment Introductory Physics Courses

ROBERT DUFRESNE, JOSE MESTRE, DAVID M. HART
AND KENNETH A. RATH
University of Massachusetts
Amherst, MA 01003 USA
dufresne@physics.umass.edu
dhart@cs.umass.edu

This study compares the effect of web-based homework (WBH) and paper-and-pencil homework (PPH) on student achievement as measured by exam performance. Various offerings of two large introductory physics sequences were examined over a three-year period, with some courses taught with PPH and some with WBH. We found that WBH offerings led to higher overall exam performance; the mean difference between WBH and PPH courses was about one-third of a typical exam standard deviation. In one WBH-PPH comparison, where there were matched final exams, the difference in exam performance was statistically significant. In addition, significant cost savings were realized in moving from PPH to WBH. Several mechanisms to explain the differential performance in favor of WBH are proposed.

Web-based homework (WBH) systems are proliferating in the teaching of large introductory physics courses[1] nationwide. With the downsizing of both science faculty and teaching assistants, many universities have abandoned time-intensive approaches to homework, such as collecting and grading paper homework and conducting small discussion sections where instructors go

over homework problems. WBH systems can provide an affordable alternative to traditional approaches to administering homework. It is likely that the financially motivated trend of replacing traditional paper-and-pencil homework (PPH) with WBH will continue to expand.

This study examines whether there is value added when PPH is replaced by WBH in large introductory physics courses. Specifically, whether there are measurable differences in exam performance between students in large introductory physics classes using WBH and students in the same introductory courses using PPH was investigated. The study investigates exam performance in two service course sequences at the University of Massachusetts at Amherst (the two-semester sequence for life-science majors and the two-semester sequence for physical science and engineering majors) over a three-year period.

The remainder of this article is divided as follows. It begins with a review of the research literature dealing with the effect of homework on achievement. Then the WBH system used in the courses investigated is described, followed by a description of the different course offerings. Results of the analysis of the effect of WBH on test performance are then presented. The conclusion is a discussion of the findings and implications.

PREVIOUS STUDIES OF THE EFFECT OF HOMEWORK
ON STUDENT PERFORMANCE

Two research strands are relevant to this study: (a) the general effect of homework on academic performance and (b) the specific effect of WBH on science/math achievement. These two areas are treated separately.

General Effect of Homework on Performance.

Although somewhat dated, several review articles have summarized studies of the relationship between homework and academic performance (Cooper, 1989; Keith, 1982; Paschal, Weinstein, & Walberg, 1984). One review that examined 15 quantitative studies (Paschal, Weinstein, & Walberg, 1984) concluded there is a moderately large positive effect of homework on achievement, especially homework that teachers grade or write comments on.

Another review (Cooper, 1989) examined 120 studies of the effect of homework on achievement. Twenty of the studies compared the achievement of students (in grades 1-12) who were given homework assignments

with students who were not. Fourteen (or 70%) produced effects favoring homework. However, the effectiveness of homework varied dramatically across grade level; high school students reaped significant benefits from homework, whereas elementary school students showed no effect (positive or negative) of homework on performance. Although the results from these studies suggested that homework can have a beneficial effect on academic performance (as measured by grades) in a variety of subject areas, there are enough counter examples to prevent stating unequivocally that doing homework improves achievement. Cooper also reviewed 50 studies that correlated achievement with the amount of time students reported spending on homework, and found that in 43 studies (or 86%), a positive correlation was found indicating that students who spent more time on homework received better grades; the remaining 7 studies indicated the opposite. As in the previous homework/no-homework comparison, the effect was nonexistent for elementary school students, and largest for high school students. A positive correlation between time spent on homework and grades was also reported in a more recent study of high achieving high school students (Tymms & Fitzgibbon, 1992).

Keith (1982) examined the effect of time spent doing homework on high school grades for a very large, representative sample of high school seniors. He found that time spent on homework was the second best predictor of grades after intellectual ability. Perhaps the most interesting finding in the study was that homework appeared to have compensatory effects; students of lower ability were able to achieve grades commensurate with higher ability students through increased time on homework. For example, Keith found that, by doing 1-3 hours of homework per week, the average low ability student was able to achieve grades commensurate with an average ability student who did not do any homework.

Effect of WBH on Performance.

Few rigorous studies on the impact of WBH on course performance exist. Most of the literature on the use of WBH in the sciences describe how courses are structured or how the WBH system itself is structured (e.g., Lee & Heyworth, 1997; Smith & Stovall, 1996; Spain, 1996; Wells & Marsh, 1997). Statements about the cognitive benefits of WBH are often anecdotal, without rigorous supporting evidence. For example, one study claimed that WBH "increases the quality of instruction" but offers no data or evidence to back up the claim (Smith & Stovall, 1996). Another claimed "while there is

no formal evidence of improved student performance, alumni have told us that they greatly benefited from use of technology in this course..." (Wells & Marsh, 1997). Another article describing a WBH system in physics, designed in large part to administer conceptual questions, claimed, "No formal evaluation of the effectiveness of these questions has been made..." (Kashy, Graff, Pawley, Stretch, & Wolfe, 1995).

Three studies offering evaluations of WBH are worth noting—two in physics courses and one in a statistics course. The statistics course (Porter & Riley, 1996) used WBH characterized as a "drill program." Students were divided into computerized homework versus noncomputerized homework groups, and performance on exams was compared. Although the computerized homework class outperformed the noncomputerized homework class, the differences were only significant in the final exam, and only on questions that related to the homework; the computerized homework class actually performed lower than the non-computerized homework class on questions not related to homework. Another study investigated the effect of multiple-choice WBH in an introductory physics setting (Lewis, Harper, & Wilson, 1991). The class was divided into two groups, and for half the semester one group experienced WBH, the other problem-solving classes (i.e., discussion sections) where an instructor went over problems in a step-by-step fashion for the class; for the second half of the semester, the two groups switched roles. Electronic homework counted 5% of the grade, but homework was not collected or graded for the problem-solving classes. Findings revealed that significantly more students did electronic homework than attended the problem-solving classes (71% attempted all electronic homework assignments vs. 12% for the problem-solving classes). At the end of the course students preferred electronic homework to problem-solving classes (65% vs. 14%). Although there were no significant differences between the groups on the mid-term exam, there was a small but significant difference (7%) in the final exam in favor of WBH. Finally, a third study investigated the effectiveness of replacing recitation sections with WBH in a large physics class (Thoennessen & Harrison, 1996). With electronic homework counting 35% of the grade, the vast majority of students completed all assignments. Findings suggested that, although doing well on the electronic homework is a necessary condition for doing well on the final exam, it is not a sufficient condition; students who did not do well on homework did not do well on the final, but doing well on homework did not correlate well with final exam performance.

OVERVIEW OF CURRENT STUDY

The current study examines several questions motivated by prior research. Primary among these is whether WBH leads to higher levels of achievement (as determined by course exam scores) than does PPH. Beyond this primary question, a number of secondary questions are considered: (a) Does WBH have a differential impact on achievement among students with high and low math ability (as determined by students' math SAT scores)? (b) Does WBH have a differential impact on achievement among high and low achievers (as determined by students' exam scores)? (c) Does WBH have a differential impact on achievement among high and low homework performers (as determined by students' homework scores)? and (d) Does WBH impact the amount of time students spend doing homework relative to PPH (as determined by student self-reporting of time spent on homework)?

This study was done *ex post facto,* and as such, is not the result of a carefully controlled experiment. Each instructor was free to make his or her own teaching decisions. They did so independent of each other, and without reference to the study; in fact, many of the classes occurred before the study was initiated. No conditions were placed on the nature, delivery, or grading of homework. Further, all exams were developed, administered, and graded by the instructor. No controls were in place to ensure that the homework treatments or exams were commensurate across different courses, or across different classes taught by the same instructor. This study does not attempt to compare a specific WBH approach with a specific PPH approach. Instead, it examines the relative impact of WBH and PPH in different course settings where the instructors made all of the teaching decisions, including the way WBH and PPH were structured.

DESCRIPTION OF WBH SYSTEM USED

The WBH system used in physics courses at the University of Massachusetts at Amherst is called Online Web-based Learning, or OWL. Originally developed in 1996 to replace an older, successful electronic homework system used in chemistry courses, OWL is currently used by more than a dozen disciplines, with more than 10,000 student users annually. OWL was adapted for use in Physics in 1997. It was first used in one introductory physics class, and has since been expanded to a dozen physics courses enrolling 2800 students/year. OWL is available 24 hours a day 7 days a week, and students appreciate and take advantage of this convenience (85% of the

roughly 60,000 OWL physics assignments done in Spring of 2000 were initiated from dorm rooms or off-campus housing).

OWL is a web-based system where students log on to do regular homework assignments. Assignments are graded automatically and the results recorded in a database. When a student submits a problem answer, OWL grades it and lets the student know whether the answer is correct (and may, at the instructor's choice, give the answer to the problem) and provides a descriptive body of feedback describing a solution strategy. The student is then allowed to retry the problem, but must work a new variation of the problem with different variable values than those used in the earlier attempt. OWL questions are parameterized, so that variable values are randomly assigned by the system each time the problem is generated. In this way students can try and retry a problem many times, getting guidance from the feedback each time a solution is submitted. The types of problems used in OWL assignments are nearly identical to those used in previous PPH assignments; in fact, the vast majority of problems in the OWL library have come from the end-of-chapter problems of standard textbooks (with permission from the publishers).

CONTEXT: GENERAL STRUCTURE OF SERVICE COURSES

The analysis described in the next section comparing the effect on test performance of WBH and traditional PPH was done on 15 offerings of four service courses at UMass-Amherst. Four different instructors taught the various offerings over a period of seven semesters, starting in the spring of 1997 and ending in the spring of 2000. Since the manner in which each course was taught was decided individually by each instructor, there were both similarities and differences across the courses.

The four service courses have UMass designations 131, 132, 151, and 152. The 131-132 courses are an algebra-based, two-semester sequence required of life-science majors, and cover most topics in the textbook (Wilson & Buffa, 1997). Typical enrollment is about 270 for 131 and 230 for 132. The 151-152 courses are a calculus-based, two-semester sequence required of those majoring in engineering and the physical sciences, and cover most topics in the textbook except modern physics and optics (Halliday, Resnick, & Walker, 1997). Typical enrollment for the spring and fall offerings of 151 are about 290 and 200, respectively; spring and fall offerings of 152 enroll about 150 and 250 students, respectively. All courses are taught in a large lecture setting.

Table 1 provides a summary of the courses analyzed in this study. The courses grouped at the top of Table 1 used traditional PPH in some form, and held weekly recitation sections led by a professor or a teaching assistant (TA), with the time in recitation usually spent answering questions that students had on homework problems.

In the fall 1997 offering of the 131 course, and fall 1997 152 course, homework was collected once a week and a TA graded some portion of it, which comprised the students' homework grade for the week. In the weekly recitation periods of the spring 1998 offering of 132, students worked in groups on a worksheet (containing 2-4 questions) designed by the instructor, and were given a grade of 0 or 1 based on effort. Homework was assigned but neither collected nor graded, although solutions were posted on the course web page, and the instructor encouraged students to come to daily office hours to go over any difficulties with homework. In the spring 1997 and spring 1998 offerings of 131, a homework set of 10 to 15 problems was assigned each week. At the end of each week the instructor would identify three problems to be collected and graded. The top 10 graded homework sets for each student were used to determine a homework grade. In the spring 1997 offering of 151, a weekly set of problems was assigned, and the recitation instructor picked one of them for the students to work out and pass in at the end of the recitation period to be graded. Students worked in groups of three on the designated problem, and could get help from the discussion instructor when they were stuck. Almost everyone who attended a recitation section in the 151 course would get a perfect homework score for the week. Overall, the total homework score was scaled and counted between 10-20% of the course grade, depending on the course.

The homework structure of the WBH offerings (the courses grouped at the bottom of Table 1) was much more uniform. Each week, students were given three assignments of between two to four problems. Students were allowed as many tries as they wished to get the problems correct. Most instructors allowed students to miss some small fraction of the homework without penalty, and the total homework score was scaled and counted between 10-20% of the course grade, depending on the instructor.

Table 1
Summary of Courses Analyzed

PAPER-AND-PENCIL HOMEWORK					
Course	Instructor	Year	Semester	Group Work During Class	SAT Cutoff
131	Professor A	1997	Fall	Yes	580/590
131	Professor B	1997	Spring	No	570/580
131	Professor B	1998	Spring	No	570/580
132	Professor C	1998	Spring	No	580/590
151	Professor D	1997	Spring	Yes	630/640
152	Professor D	1997	Fall	Yes	630/640
WEB-BASED HOMEWORK					
Course	Instructor	Year	Semester	Group Work During Class	SAT Cutoff
131	Professor A	1998	Fall	Yes	580/590
131	Professor A	1999	Fall	Yes	580/590
131	Professor B	1999	Spring	No	570/580
131	Professor B	2000	Spring	No	570/580
132	Professor C	1999	Spring	No	580/590
151	Professor D	1998	Spring	Yes	630/640
151	Professor D	1999	Spring	Yes	630/640
152	Professor D	1998	Fall	Yes	630/640
152	Professor D	1999	Fall	Yes	630/640

There were similarities and differences across the courses in how the three weekly lectures were conducted. The column labeled "group work during class" in Table 1 denotes whether or not the instructor employed some form of active learning during class, which generally consisted of having students work collaboratively in groups to solve problems. In all offerings by Professor D and Professor A the *Classtalk[2]* classroom communication system was used to facilitate the collection and display of student answers during collaborative group problem solving. The most "radical" offerings were those of Professor D, who did very little lecturing on the material in the textbook. He expected students to come to class having read the textbook, and class time was used to refine students' understandings and to help them apply the knowledge they acquired from the textbook to solve problems. In Professor D's offerings, students would spend most of the class time working collaboratively on problems, the answers to which were submitted electronically for

display in histogram form to the entire class. A subsequent class-wide discussion aired the solution strategies leading to the answers displayed in the histogram (i.e., individual students volunteered and argued for their solution strategy, with the rest of the students listening and evaluating the arguments presented).[3] In Professors A's offerings, he lectured on the material in the textbook, with short periods of collaborative problem solving interspersed during the lecture. Typically the collaborative problem solving would deal with material that was just covered during lecture. No class-wide discussion of the solution strategies ensued; rather, after students submitted their answers, Professor A discussed the solution to the problem that the students worked on in groups.[4]

DATA SOURCES AND TREATMENT GROUPS

Data Sources

For each class, three sources of data were analyzed: (a) exam scores, (b) homework scores, and (c) math SAT scores.

Exam scores—All classes administered four exams over the course of the semester. Each exam was graded on a scale of 100 points. Each student was assigned an exam score equal to the average of his or her four class exams. The students' average exam scores were used to compute a class average and standard deviation. Generally, exam performance between classes on a single exam are not compared, since there is no way to control for exam content or level of exam difficulty. Professor D's 151 classes are an exception. The final exams for the three different classes contain a subset of identical questions. To analyze the 151 final exam from different semesters, each student was assigned a score based on the proportion correct out of the total number of identical questions across exams. The final exam scores were scaled to be out of a total of 100 points to make the presentation of results consistent with the presentation of other exam data.

Exam scores are used in two different ways. Their primary use is as a measure of course achievement, but they are also used to divide students in each class into two groups based on achievement level. These groups are used to determine whether there were differential effects of WBH related to achievement. For each class, an exam cutoff was chosen. Students with scores below the cutoff were placed in the Low Exam group and students with scores above the cutoff were placed in the High Exam group. For each

class the cutoff was chosen to make the number of students in each group as equal as possible.

Math SAT scores—Math SAT scores were acquired through the university for nearly all of the students participating in the study. Math SAT scores were chosen as an indicator of mathematical ability. To use this variable, all students associated with a given instructor-course combination were divided into two groups based on a SAT cutoff. Students with SAT scores below the chosen cutoff were placed in the Low SAT group. Students above the cutoff were placed in the High SAT group. In each case the cutoff was chosen to make the number of students in each group as equal as possible. The cutoffs can be found in Table 1.

Homework scores—Homework scores varied in scale across classes, even in those classes taught by the same instructor. Using the same procedure used to divide classes into groups based on Exam and SAT scores, each class was divided into two roughly equal groups based on homework scores: a High Homework group and a Low Homework group. The researchers were not always able to achieve nearly equal groups. An extreme example of this was Professor A's 131 PPH class, where close to 80% of the students got full credit on the homework, causing the high homework group to greatly outnumber the low homework group in that class. In most cases, the groups differed in size by a few percentage points. There were no significant differences between the mean SAT scores of the different homework groups.

Treatment Groups

For the purpose of analysis two treatment groups were distinguished: (a) those students receiving PPH and (b) those students receiving WBH. It was generally the case that for a given instructor-course combination, the classes receiving PPH occurred prior to the classes receiving WBH. In all cases the WBH was delivered using OWL.

RESULTS

For each subsection first the results for all courses are reported, followed by the results of the matched set of questions from Professor D's final exams.

WBH Versus PPH Overall

In this section results relevant to the impact of WBH on student achievement in physics as measured by exam performance are presented.

Results for all service courses—Table 2 contains the mean exam scores for all participating classes. Each instructor-course combination occurred at least once with PPH and once with WBH. For those instructor-course combinations where there were two PPH (WBH) classes, the mean exam scores for the two classes were averaged to obtain a mean PPH (WBH) score. A gain was computed for each instructor-course combination by taking the difference between the mean WBH score and the mean PPH score. Four of the five instructor-course combinations show positive gains ranging from 4.9 to 9.7. One instructor-course combination shows a small negative gain of –2.2. The mean gain for the five instructor-course combinations is 4.96. This difference is not statistically significant ($t(4) = 2.485$, $p = .068$) and is roughly a third of a typical class standard deviation. (It should be noted that if it could be assumed that the PPH and WBH classes for a given instructor-course combination were equivalent, except for type of homework, then PPH and WBH classes could be compared directly for each instructor-course combination. A direct comparison of this kind shows that the four positive gains are statistically significant and the one negative gain is not.)

Table 2
Mean Exam Scores for All Service Courses
(For individual classes, the first number represents the mean exam score [out of 100 points], followed by the standard deviation in parentheses and the number of students in that class in brackets)

Instructor-Course	PPH 1	PPH 2	PPH Average	WBH 1	WBH 2	WBH Average	Gain
Professor A131	63.1 (15.7) [311]		63.1 (14.6) [190]	71.9 (14.4) [212]	64.0	68.0	4.9
Professor B 131	62.5 (17.8) [78]	65.0 (13.4) [68]	63.8 (14.1) [76]	57.2 (16.2) [32]	66.0	61.6	-2.2
Professor C 132	53.8 (14.3) [159		53.8 (15.2) [162]	58.8		58.8	5.0
Professor D 151	50.0 (16.8) [181]		50.0 (16.6) [253]	56.1 (12.8) [262]	58.7	57.4	7.4
Professor D 152	53.1 (14.0) [181]		53.1 (12.2) [207]	61.6 (13.0) [193]	63.9	62.8	9.7

Results for Professor D's final—The final exams given in Professor D's 151 classes contain 17 questions in common. This situation permits the most direct comparison of the impact of WBH and PPH on achievement. Table 3 contains the mean final-exam score for Professor D's 151 classes. There are three classes all together, one PPH class and two WBH classes. The mean final-exam score for the two WBH classes are higher than the mean final-exam score for the PPH class. The 7.3 point difference between the WBH 1 and PPH classes is statistically significant ($t(449) = 4.262$, $p = 0.001$). However, the 2.4 point difference between the WBH 2 and PPH classes is not statistically significant.

Table 3

Mean Final Exam Scores for Professor D's Classes

	Mean	SD	N
PPH	45.7	16.7	171
WBH 1	53.0	18.7	231
WBH 2	48.1	16.4	266

WBH Versus PPH By SAT Group

In this section results relevant to the differential impact of WBH on student achievement in physics for students with different math ability as measured by math SAT score is presented

Results for all service courses—All students associated with a given instructor-course combination were divided into two groups based on an SAT cutoff. Mean PPH and WBH exam scores were then computed for each SAT group. Mean PPH and WBH exam scores (averaged across all instructor-course combinations) are shown by SAT group in Figure 1. Not unexpectedly, the High SAT group outperformed the Low SAT group. The average exam score for the PPH-High SAT group is 14.3 points above the average exam score for the PPH-Low SAT group. Similarly, the average exam score for the WBH-High SAT group is 12.4 points above the WBH-Low SAT group. Both differences are statistically significant. The mean gain (i.e., difference between the mean WBH exam score and the mean PPH exam score) for the Low SAT group (6.1) is slightly larger than the mean gain for the High SAT group (4.2). However, this differential gain between High and Low SAT groups is not statistically significant. Nevertheless, four of the five instructor-course combinations show the same pattern of differential gain.
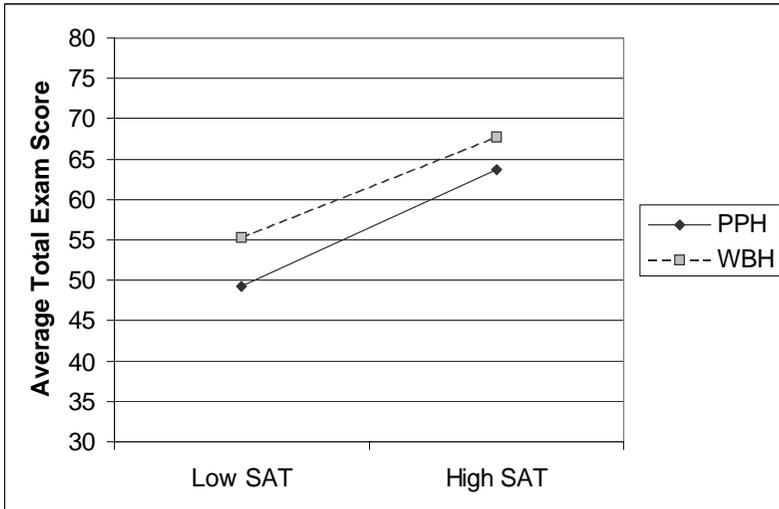
Figure 1. Average exam scores for all classes across Math SAT groups

Results for Professor D's final exam—The mean PPH and WBH final
exam scores for Professor D's classes are shown by SAT group in Figure 2.
Similar to the results for all service courses, the High SAT group outper-
formed the low SAT group by 10 points (PPH class), 11.3 points (WBH 1
class), and 9.4 points (WBH 2 class). All differences are statistically signifi-
cant. When comparing the PPH class with the WBH 1 class, the mean gain
(i.e., difference between the mean WBH 1 exam score and the mean PPH
exam score) for the Low SAT group (7.2) was slightly lower than the mean
gain for the High SAT group (8.0). When comparing the PPH class with the
WBH 2 class, the mean gain (i.e., difference between the mean WBH 2
exam score and the mean PPH exam score) for the Low SAT group (3.5)
was slightly higher than the mean gain for the High SAT group (1.2). The
differential gain between High and Low SAT groups is not statistically sig-
nificant. Further, the pattern of differential gains was not consistent for the
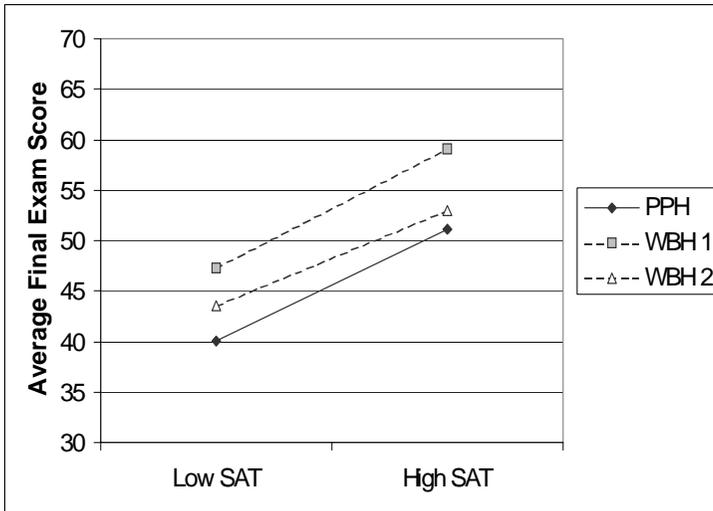two WBH classes.

Figure 2. Average exam scores for Professor D's classes across Math SAT groups

WBH versus PPH By Exam Group

In this section results relevant to the differential impact of WBH on student achievement in physics for students with different levels of achievement as measured by exam score are presented.

Results for all service courses—All students associated with a given instructor-course combination were divided into two groups based on exam score. Mean PPH and WBH exam scores were then computed for each exam group. Mean PPH and WBH exam scores (averaged across all instructor-course combinations) are shown by exam group in Figure 3. As a consequence of the "exam group" definition, the High Exam group outperformed the Low Exam group. The average exam score for the PPH-High Exam group is 23.8 points above the average exam score for the PPH-Low Exam group. Similarly, the average exam scores for the WBH-High Exam group are 23.1 points above the PPH-Low Exam group. These differences are large and statistically significant. The mean gain (i.e., the difference between the mean WBH exam score and the mean PPH exam score) for the Low Exam group (5.3) is slightly larger than the mean gain for the High

Exam group (4.6). This differential gain between High and Low Exam groups is not statistically significant. Nevertheless, four of the five instructor-course combinations show the same pattern of differential gain.
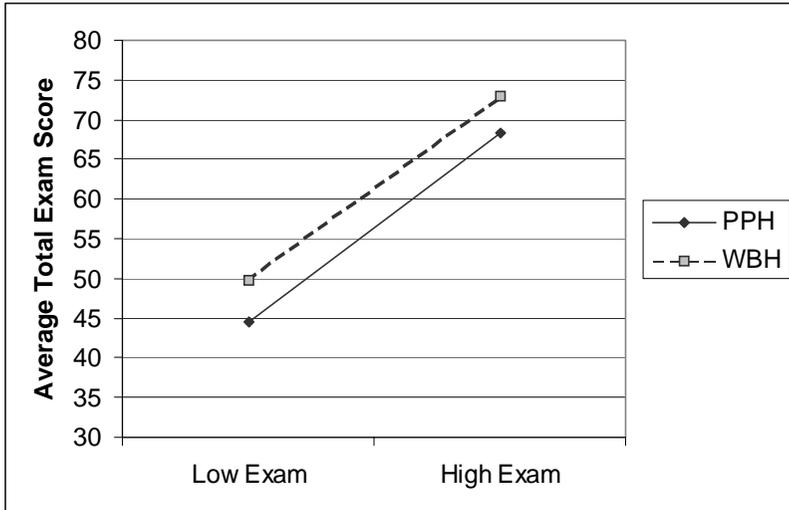

Figure 3. Average exam scores for all classes across Exam groups

Results for Professor D's final exam—The PPH and WBH final exam scores for Professor D's classes are shown by exam group in Figure 4. The High Exam group outperformed the Low Exam group by 26.9 points (PPH class), 29.5 points (WBH 1 class), and 26.9 points (WBH 2 class). All differences are large and statistically significant. When comparing the PPH class with the WBH 1 class, the mean gain (i.e., difference between the mean WBH 1 exam score and the mean PPH exam score) for the Low Exam group (4.9) was lower than the mean gain for the High Exam group (7.5). This differential gain between High and Low exam groups is not statistically significant. When comparing the PPH class with the WBH 2 class, the mean gain (i.e., difference between the mean WBH 2 exam score and the mean PPH exam score) for the Low Exam group (1.0) was essentially the same as the mean gain for the High Exam group (1.0).
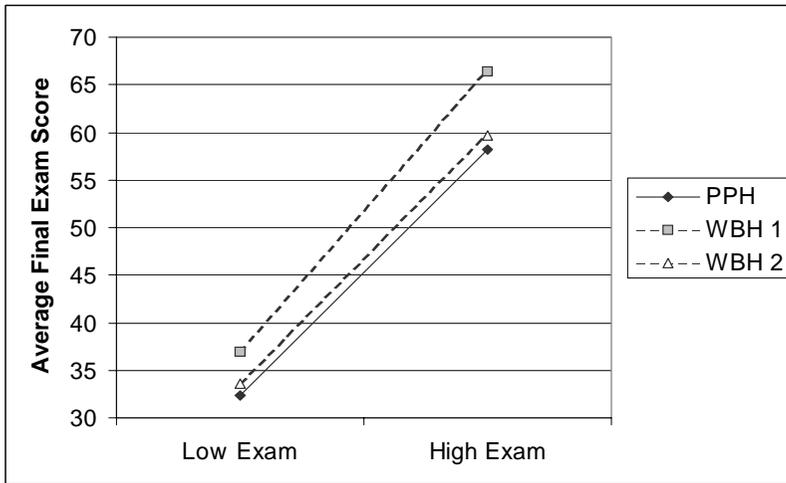
Figure 4. Average exam scores for Professor D's classes across Exam groups

WBH versus PPH By Homework Group

In this section results relevant to the differential impact of WBH on student achievement in physics for students with different levels of homework performance are presented.

Results for all service courses—All students associated with a given instructor-course combination were divided into two groups based on homework score. It was not always possible to achieve (approximately) equal sample size for the two homework groups. Mean PPH and WBH exam scores (averaged across all instructor-course combinations) are broken down by homework group in Figure 5. The mean exam score for the PPH-High Homework group is 9.1 points above the mean exam score for the PPH-Low Homework group. Similarly, the average exam scores for the WBH-High Homework group are 7.1 points above the WBH-Low Homework group. These differences are statistically significant. The mean gain (i.e., the difference between the mean WBH exam score and the mean PPH exam score) for the Low Homework group (7.4) is larger than the mean gain for the High homework group (5.5). This differential gain between High and Low Homework groups is not statistically significant. Only three of the five instructor-course combinations show the same pattern of differential gain.

Results for Professor D's final exam—The PPH and WBH mean final exam scores for Professor D's classes are broken down by homework group in Figure 6. The High Homework group outperformed the Low Homework group by 5.9 points (PPH class), 9.2 points (WBH 1 class) and 6.7 points (WBH 2 class). All differences are statistically significant. When comparing the PPH class with the WBH 1 class, the mean gain (i.e., difference between the mean WBH 1 exam score and the mean PPH exam score) for the Low Homework group (7.3) was lower than the mean gain for the High Homework group (10.6). When comparing the PPH class with the WBH 2 class, the mean gain (i.e., difference between the mean WBH 2 exam score and the mean PPH exam score) for the Low Homework group (3.8) was slightly smaller than the mean gain for the High Homework group (4.6). The differential gain between High and Low Homework groups is not statistically significant.
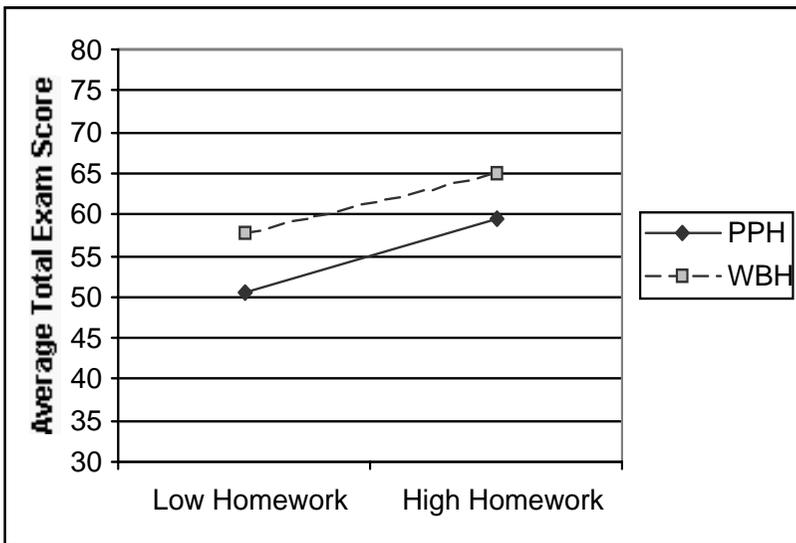

Figure 5. Average exam scores for all classes across Homework groups

Figure 6. Average final exam scores for Professor D's classes across homework groups

WBH Versus PPH –Time on Task

For Professor D's PPH and WBH 1 classes, data on time spent doing homework was collected as part of the course. The data are the result of student self-reporting and were collected near the end of the semester as part of a larger survey. These were the only classes for which data on time spent doing homework was collected. The results are presented in Table 4. The information is included here to give the reader a sense of the possible variation in time-on-task doing homework in the different classes. In this instance, the majority of the PPH students report spending less than one hour doing homework each week. In contrast 46% of WBH students spent more than four hours, and another 43% spent between two and four hours. Whether students spent more time doing WBH than PPH in the other courses is unknown.

Table 4

Time Students Reported Spending on Homework in Professor D's Class for
Paper and Pencil (PPH) Homework versus the First Semester
of Web-Based Homework (WBH 1)

	WBH 1		PPH	
	#Responses	%Responses	#Responses	%Responses
Less Than Two Hours	23	11.0%	162	62.1%
Two To Four Hours	90	42.9%	88	33.7%
More Than Four Hours	97	46.2%	11	4.2%

DISCUSSION AND INSTRUCTIONAL IMPLICATIONS

Overall, it was found that replacing PPH with WBH led to higher physics achievement as measured by exam performance. The difference in performance was about a third of a typical exam standard deviation for a given class. Further, statistical significance could only be established in the courses taught by Professor D, where a portion of each final exam was the same across three different offerings of the same course. The findings indicate that students who perform well on homework generally attain higher achievement scores**.** However, no consistent evidence was found of a differential impact when replacing PPH with WBH; that is, the performance gaps in exam scores between WBH and PPH were similar for low and high SAT-math groups, for low and high homework score groups, and for low and high exam performance groups.[5]

Possible explanations for the better overall performance on exams for WBH classes are now offered. The student self-report data on time spent on homework suggests that students spend more time doing WBH than PPH. Further, all of the WBH was graded, with students receiving immediate feedback, as well as hints following the submission of an incorrect answer to help them head in the right direction; in contrast, only some of the PPH was graded and it was often the case that students could get maximum credit by doing minimal homework. The research literature reviewed earlier indicates not only that time spent on homework was the second best predictor of grades after intellectual ability, but also that homework is most effective when teachers grade and write comments on it.

Some caveats are in order to place this study in the proper perspective. The authors had no way of controlling for exam difficulty from one year to the next, except for the final exam in one course where the instructor used the same subset of questions in three consecutive years. Differences between any two classes could be the result of differences in exam difficulty. Consequently, the authors had to treat the aggregate exam performance of each course as a single data point, as opposed to treating each individual student's exam performance as a data point. Additionally, all instances of WBH followed offerings with PPH. Consequently the authors were not able to account for any possible changes in exam performance resulting from the order in which the PPH and WBH courses were offered.

Neither the PPH nor the WBH used in the analyses were designed to optimize learning. The types of questions used in WBH were quite similar to those used in PPH, and consisted of end-of-chapter problems. Although WBH provided some feedback in the form of suggested solution strategies when students submitted a wrong answer, this feedback was not tailored to meet students' specific needs but rather it was based on the instructor's perception of what students might find useful for solving the problems. Nevertheless, it is easy to imagine how future WBH systems could be designed to provide individualized instruction similar to that of a human tutor at a fraction of the cost. For example, future WBH systems could contain artificial intelligence features to help diagnose and treat students' specific difficulties, as well as features that tailor feedback to match students' learning styles.

This study suggests that even the most straightforward adoption of WBH is at least comparable to PPH in terms of promoting student achievement. This is an important finding in view of the cost benefits associated with computer-administered homework in large introductory science courses. A careful cost benefit analysis of the introduction of WBH in the Department of Physics at UMass indicated savings in excess of $130,000 annually in faculty and (TA) resources because small discussion sections and hand-graded homework assignments have been eliminated in courses using WBH (Rath & Hart, in preparation). These types of savings are critical at a time when most large universities have been downsizing both faculty and TA lines without an accompanying reduction in teaching loads. In short, there are additional benefits in shifting from PPH to WBH.

Thus far most applications of WBH in large introductory courses are simple attempts to move traditional homework into electronic form. Given the proliferation of WBH, answers to the following important research questions would inform how we could structure the experience to maximize learning for students: (a) Are there differences in test performance between

WBH and PPH if time spent doing homework were controlled for? (b) Can WBH systems be designed that actively promote the development of expert skills, such as analyzing problems qualitatively based on principles and concepts, structuring one's knowledge based on a hierarchy of principles/concepts, and becoming self-reflective about one's own learning? (c) How easily and cost-effectively could relevant findings from research into intelligent tutoring systems be applied in the design of WBH systems to maximize learning? and (d) How easily and cost-effectively could diagnostic features be incorporated into WBH systems to identify different student learning styles so that feedback is matched to students' learning styles? Although our findings suggest that students appear willing to spend more time doing WBH than PPH, research on learning also suggests that simply throwing more time at a task is not necessarily the most efficient way of learning a subject as complex as science, especially if students are practicing novice-like strategies to solve problems, such as equation manipulation (National Research Council, 1999).

References

Cooper, H. (1989). Synthesis of research on homework. *Educational Leadership, 47*(3), 85-91.

Dufresne, R.J., Gerace, W.J., Leonard, W.J., Mestre, J.P., & Wenk, L. (1996). *Classtalk*: A classroom communication system for active learning. *Journal of Computing in Higher Education, 7*(2), 3-47.

Halliday, D., Resnick, R., & Walker, J. (1997). *Fundamentals of physics* (5th ed.). New York: Wiley.

Kashy, E., Graff, S.J., Pawley, H., Stretch, W.L., & Wolfe, S.L. (1995). Conceptual questions in computer-assisted assignments. *American Journal of Physics, 63*(11), 1000–1005.

Keith, T.Z. (1982). Time spent on homework and high school grades: A large sample path analysis. *Journal of Educational Psychology, 74*(2), 248-253.

Lee, F., & Heyworth, R. (1997). *Electronic homework*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED405859)

Lewis, R.A., Harper, B.M., & Wilson, M. (1991). Computer assignments and problems classes for physics students. *Computers and Education, 16*(4), 349–362.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

Meltzer, D.E., & Manivannan, K. (1996). Promoting interactivity in physics lecture classes. *The Physics Teacher, 34*(2), 72-76.

Mestre, J.P., Gerace, W.J., Dufresne, R.J., & Leonard, W.J. (1997). Promoting active learning in large classes using a classroom communication system. In E.F. Redish & J.S. Rigden (Eds.), *Proceedings of International Conference on Undergraduate Physics Education* (pp. 1019-1036). Woodbury, NY: American Institute of Physics.

National Research Council (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Paschal, R.A., Weinstein, T., & Walberg, H.J. (1984). The effects of homework on learning: A quantitative synthesis. *Journal of Educational Research, 78*(2), 97-104.

Porter, T.S., & Riley, T.M. (1996). The effectiveness of computer exercises in introductory statistics. *Journal of Economic Education, 27*(4), 291–299.

Rath, K.A., & Hart, D.M. (2002). Online web-based learning (OWL): Systematic integration of online homework into undergraduate courses. Department of Computer Science, University of Massachusetts, Amherst. Retrieved from the World Wide Web June 11, 2002, from: http://ccbit.cs.umass.edu/owl/

Smith, S., & Stovall, I. (1996). Networked instructional chemistry: Using technology to teach chemistry. *Journal of Chemical Education, 73*(10), 911–915.

Spain, J.D. (1996). Electronic homework: Computer-interactive problem sets for general chemistry. *Journal of Chemical Education, 73*(3), 222–225.

Thoennessen, M., & Harrison, M.J. (1996). Computer-assisted assignments in a large physics class. *Computers in Education, 27*(2), 141–147.

Tymms, P.B., & Fitzgibbon, C.T. (1992). The relationship between homework to A-level results. *Educational Research, 34*(1), 3-10

Wells, K.L., & Marsh, L.C. (1997). Key aspects of a computerized statistics course. *Journal of Computing in Higher Education, 8*(2), 72–93.

Wilson, J., & Buffa, A. (1997). *College Physics* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Note

Acknowledgments

Notes

1. By far the most widely used stand-alone WBH system is WebAssign, marketed out of North Carolina State University; the University of Massachusetts–Amherst developed and disseminates the Online Web-based Learning system (OWL); Michigan State University developed and disseminates the CAPA system; the University of Illinois uses both CyberProf™and Tycho. In addition, there are many users of the web-based homework components of on-line management systems, such as Web-CT, Blackboard, and others.
2. Classtalk is marketed by Better Education, Inc., Yorktown, VA.
3. For more details on this instructional method, see Dufresne, Gerace, Leonard, Mestre, and Wenk, 1996; Mestre, Gerace, Dufresne, and Leonard, 1997.
4. This approach is similar to Mazur's (1997) *Peer Instruction*. In Professor A's offerings, students would use flashcards to convey their answers, as described in Meltzer and Manivannian, (1996).
5. We also looked for evidence of compensatory effects, namely that the performance gap between high ability students doing no homework and low ability students doing homework would change as a function of replacing PPH by WBH, but no such evidence surfaced.