International Journal of Engineering Education Vol. 27, No. 6, pp. 1333–1342, 2011
Printed in Great Britain

0949-149X/91 $3.00+0.00

Do Homework Grading Policies Affect Student Learning?*

ALI YALCIN
University of South Florida, Industrial and Management Systems Engineering Department, 4202 East Fowler Ave., ENB 118, Tampa, FL 33620 USA. E-mail: ayalcin@usf.edu

AUTAR KAW
University of South Florida, Mechanical Engineering Department, 4202 East Fowler Ave., ENB 118, Tampa, FL 33620 USA. E-mail: kaw@usf.edu

A significant amount of teaching assistant and instructor time is allocated to grading homework assignments, especially in large enrolment courses. However, the benefits of such a time- consuming activity are unknown and not well documented. Our goal is to examine the impact that different homework grading policies have on students' final examination performance. We are interested not only in the overall student performance, but also in the performance of specific student subgroups with varying backgrounds, as well as the impact of homework on the type of learning that takes place in the course. The study was conducted in a Numerical Methods course at the University of South Florida over a period of three years encompassing data from over 300 Mechanical Engineering students. Statistical analysis of data regarding the impact of homework grading policies on student subgroups based on several factors is presented. Our results indicate that there is no statistically significant difference in student examination performance when homework is graded versus when homework is assigned but not graded. However, certain grading policies did seem to put some subgroups of students at a disadvantage. While grading homework may not be critical in improving student examination performance, it is important to ensure that students practice the concepts.

Keywords: homework; grading policy; student performance

1. Introduction

The assumption that assigned homework contributes to effective learning in higher education is a commonly accepted premise [1–4]. Probably the most well-known and sustained research into the benefits of homework for the general student population including K-12 students is by Cooper and associates [1–2]. Their research between 1987 and 2003 involved 60 research universities and concludes that homework has a positive effect on student achievement. However, studies do not consistently show the benefits of grading homework assignments in improving examination performance, especially for students in higher education. To answer such questions categorically is becoming increasingly important as class sizes are becoming larger, while instructor and teaching assistant resources are becoming scarcer.

Weems [5] examined the effects of homework collection on achievement of college-level Intermediate Algebra students. They addressed the question of whether a student who is required to hand in homework performs better than one who does not. The study, conducted with over 108 students, was inconclusive based on the overall course grade distribution.

A more in-depth study reported by Trussell and Dietz [6] involved a sophomore-level electrical engineering course. Two sections were compared during two consecutive semesters: in one section, assigned homework was graded and, in the other section, the same homework was assigned but not graded. The section with graded homework performed significantly better in test scores in the first semester. However, there was no significant difference between the two sections in the second semester.

Peters et al. [7] studied the impact of homework on an introductory productions and operations course. Two treatments were selected for the same course. In one treatment, the instructor assigned homework from the back-of-book exercises with set due dates, graded the assignment, and returned it to the student. About 15% of the final grade was assigned to homework, while the rest was based on examinations. In the second treatment, the instructor assigned the homework just like in the first treatment, but did not collect the homework for grade. The student's final grade was based only on the examinations. They found that homework had some positive impact but it was limited to the quantitative part of the examinations.

Gutarts and Nain [8] investigated the benefits of mandatory homework in college level calculus courses. For one group, homework was collected and graded. For the other group, the same homework was assigned but not collected or graded and the students were given weekly quizzes. No significant difference was found in the examination performance of the two groups.

* Accepted 17 July 2011.

An alternative for courses that have high student enrollment and a limited number of teaching assistants are online homework systems. In recent years, this alternative has become widely accessible and used, especially at college level. An example of such a system designed specifically for assigning and grading homework is the WebAssign (http://www.webassign.net). These systems may be a viable solution but there are costs associated with the use of such systems, both for students and faculty and, at this time, availability of engineering topics in these systems is very limited. More information regarding these systems is reported in [9–11].

In this paper, we are not looking at the importance of assigning homework but at the grading of the homework, because students must practice solving problems to successfully prepare for tests and other components of a grading scheme. We are posing the following research questions:

- Does the way that homework is assigned for grading, so as to encourage good principles of teaching [12] of time-on-task and rapid feedback, improve examination performance?
- More importantly and different from other studies, what is the impact of grading homework on specific student subgroups based on gender, age, students' transfer status, and performance in prerequisite courses? While studies may show the inconclusiveness on overall examination performance, we may be neglecting the impact on certain subgroups in the process.
- Which subgroups are more likely to hand in homework?
- What is the impact of homework on the level of learning as measured by the examination that is based on Bloom's Taxonomy [13]?

The rest of the paper is organized as follows. In Section 2, we discuss the experimental design. Section 3 presents the results of the statistical data analysis in three main categories as follows:

1. the tendency to turn in homework and student performance measured by the homework grade,
2. the impact of homework grading policy on student performance and, finally,
3. how grading policy impacts the type of learning.

In Section 4, we summarize the conclusions of the study.

2. Experimental design

The study was conducted in a Numerical Methods course at the University of South Florida. The course was taught in Spring (16 weeks) and Summer (10 weeks) semesters each year. The course included the following components.

2.1 Homework

There were 20–25 homework assignments in the course. The homework assignments were the end-of-chapter exercises in the course textbook (co-authored by the second author) that reinforce the methods learned in the class. Each end-of-chapter exercise set had six multiple-choice questions based on high and low level of Bloom's taxonomy and about 6–8 free-response problems. All the end-of-chapter exercises were expected to be solved by the students in order to prepare for the exams. When homework was graded, it constituted 10–12% of the final course grade.

2.2 Computer programming projects

Projects were about interpreting data from engineering experiments and using MATLAB programming to do so. The students had to evaluate and synthesize what they have learned in class in order to develop and solve mathematical models of scientific and engineering problems. The programming concepts reinforced by the computer programming projects were not tested in the final examination. Projects made up 10–15% of the final course grade.

2.3 Examinations

The examinations were held in class and were closed book–closed notes. The examinations were a mix of questions that were multiple-choice, fill-in-the-blank, but mostly free-response where the student had to apply numerical methods to solve a mathematical procedure. The examinations were graded and returned to the student in the next class period. The examinations made up 50–53% of the final course grade.

2.4 Final examination

The final examination was used to measure the performance of the students in this study. It was a 32 multiple-choice question test based on Bloom's taxonomy. Four questions from each of the eight topics of the course comprised the examination. Two of the four questions on each topic were at the low level of Bloom's taxonomy and two were at the high level of Bloom's taxonomy. The end-of-chapter exercises assigned as homework were aligned with the final examination as half of the end-of-chapter exercise set was multiple-choice questions based on Bloom's taxonomy and so was the final examination. The final examination made up 28–36% of the final course grade.

In this study, three homework assignment and grading policies were considered as follows:

1. OPOGP (one problem assigned for collection—one problem graded policy): In summer 2006 and spring 2007, at the end of almost every class, we assigned one problem for collection and required the students to submit their solution at the beginning of the next class. Since we strongly believe in 'the good practice of giving prompt feedback' [8], we gave the graded homework back in the following class.
2. TPOGP (three problems assigned for collection—one problem graded policy): In summer 2007 and spring 2008 semesters, at the end of almost every class, we assigned three problems for collection and required the students to submit their solutions at the beginning of the next class. Only one randomly selected problem (same one for all students) was graded out of the three problems. The students' performances on the other two problems were not noted. Again, we gave the graded homework back in the following class.
3. NPNGP (no problems assigned for collection—none graded policy): In summer 2008 and spring 2009 semesters, the students were expected to do all the end-of-chapter exercise sets as in the previous treatments, but were not required to turn in any solutions. The performance of the students or the degree of completion related to these problems is not known as no work was turned in by the students. We posted solutions to only the multiple-choice questions, which comprised of about 50% of the end-of-chapter exercises, and used personal response systems in the classroom for reviewing pre-requisite knowledge and checking end-of-topic knowledge through sets of 5–10 multiple-choice questions. Since there were eight topics in the course, personal response systems were used 16 times in the semester.

Throughout the rest of the paper, these three homework grading policies will be referred to as OPOGP, TPOGP and NPNGP, respectively. Note that the same final examination was used in all six semesters, the same homework problems were assigned and graded (when applicable) during the two semesters within each grading policy, and other components of the course, such as computer programming projects and examinations, remained unchanged throughout the duration of the study.

At the end of each semester, the students' information regarding homework submissions, homework grades and final examination performance was matched with their gender, age, transfer status and prerequisite grade point average information from their student files. The data used in the study is highly reliable as all of it is drawn from official student records.

We considered several subgroups for these four factors:

1. Based on *gender*, students were separated into *male* and *female* subgroups.
2. Based on *age*, students were separated into two subgroups as *non-adult students* whose age was less than or equal to 22 at the beginning of the course and as *adult students* whose ages were greater than 22.
3. The *transfer status* indicates whether the student was a *first time college student (FTIC)*, *transferred from a community college (CC)* with a completed Associates of Arts degree, and *other (OT)* which included students who transfer from other institutions without a completed degree.
4. Finally, students were grouped by their academic performance in the course prerequisites (Calculus I, II, III, and Ordinary Differential Equations). The students were separated into two groups based on their *grade point average in the prerequisite courses (PR–GPA)* as those with *PR–GPA* < 3.0 and *PR–GPA* $\geq$ 3.0.

Table 1 shows the sample sizes and final examination score statistics for the six semesters and the grading policies considered in this study.

3. Results and discussion

3.1 Who turns in homework?

During the four course offerings between summer 2006 and spring 2007, the homework assigned for collection was graded. The data collected from

Table 1. Final examination score data (maximum possible score is 32)

Grading policy	Semester	Number of students	Average final exam. score	SD final exam. score
OPOGP	Summer 2006	56	20.1	4.75
	Spring 2007	55	21.1	3.97
TPOGP	Summer 2007	70	21.9	4.30
	Spring 2008	41	22.5	3.49
NPNGP	Summer 2008	55	21.6	3.75
	Spring 2009	41	22.0	4.42

students' homework allowed us to ask and answer some interesting questions: specifically, do certain subgroups of students have a higher tendency to turn in homework? To answer this question, we conducted an analysis of variance (ANOVA) [14] within each factor to determine if a subgroup was more likely to turn in homework than another. The results of the analysis shown in Table 2 clearly state that no subgroup based on gender, age, transfer status and, probably most surprising, the PR–GPA factors had a significantly higher tendency to turn in homework. Students in all subgroups on average turned in 88–92% of the homework assigned for collection. The results based on PR–GPA may be counterintuitive in the sense that students with a lower GPA (PR–GPA in this study) are, more often than not, thought of as those who do not do their homework.

3.2 Who turns in better quality homework?

When it comes to homework grades, the results were a little different. A similar ANOVA was conducted between the subgroups in each factor and the results are summarized in Table 3. While the subgroups within age, gender and transfer status factors showed no significant difference in students' homework performance, it is very clear that students with PR–GPA $\geq$ 3.0 outperformed students with PR–GPA < 3.0 by more than 10% on average.

3.3 Who does better in exams?

The analysis of the results in this section includes data from additional two semesters of course offerings during Summer 2008 and Spring 2009 where the students were assigned but not required to turn in homework to be graded.

The analysis of final examination performance proved to be even more interesting. The results of a similar analysis are shown in Table 4. As we have seen in homework performance analysis, students with PR–GPA $\geq$ 3.0 significantly outperformed students with PR–GPA < 3.0 in examination performance. This suggests that homework performance and not the amount of homework turned in is more closely correlated with better examination performance. In this study, the correlation analysis [15] between homework grade and final examination grade resulted in $r = 0.28$ (moderate relationship), closely following similar analysis reported elsewhere [16–17]. On the other hand, the correlation between the percentage of homework assignments turned in and the final examination performance was not present at $r = 0.09$ (weak relationship).

An unexpected finding in this analysis was related to the comparison between adult and non-adult students. While adult students were as likely to hand in homework as non-adult students and there was not a significant difference in the homework performance of the two groups, adult students

Table 2. Analysis summary of percentage of homework turned in

	Subgroup	Number of students	Avg. percentage of homework handed in	ANOVA
Gender	Male	199	90.7	$F(1,220) = 0.131, p = 0.72$
	Female	23	89.4	
Age	Adult	117	90.8	$F(1,220) = 0.107, p = 0.75$
	Non-adult	105	90.2	
Transfer status	CC	69	92.7	$F(2,219) = 1.457, p = 0.24$
	FTIC	107	88.7	
	OT	46	91.4	
PR–GPA	< 3.0	125	89.3	$F(1,220) = 1.806, p = 0.18$
	$\geq$ 3.0	97	92.1	

Table 3. Analysis summary of homework performance

	Subgroup	Number of students	Avg. homework grade (%)	ANOVA
Gender	Male	199	78.4	$F(1,220) = 0.276, p = 0.60$
	Female	23	80.4	
Age	Adult	117	77.3	$F(1,220) = 1.200, p = 0.28$
	Non-adult	105	79.9	
Transfer status	CC	69	78.6	$F(2,219) = 0.403, p = 0.69$
	FTIC	107	77.7	
	OT	46	80.5	
PR–GPA	<3.0	125	74.8	$F(1,220) = 13.259, p < 0.01$
	$\geq$3.0	97	83.3	

Table 4. Analysis summary of final examination performance

	Subgroup	Number of students	Avg. final examination grade	ANOVA
Gender	Male	284	17.2	F(1,316) = 1.569, p = 0.21
	Female	34	22.3	
Age	Adult	162	20.7	F(1,316) = 11.410, p < 0.01
	Non-adult	156	22.3	
Transfer status	CC	100	20.6	F(2,315) = 4.352, p = 0.14
	FTIC	148	21.7	
	OT	70	22.4	
PR–GPA	<3.0	174	20.1	F(1,316) = 49.903, p < 0.01
	≥3.0	144	23.2	

did not perform as well as non-adult students in the final examination.

3.4 Does grading homework matter?

The analysis of the final examination scores from the Summer and Spring semesters where the same homework grading policy was used showed no significant differences between the two semesters, which allows us to combine the data from the Spring and Summer semesters. This results in 111 students for the OPOGP, 111 students for the TPOGP, and 96 students for the NPNGP [(OPOGP: F(1,109) = 1.145, p = 0.23); (TPOGP: F(1,109) = 0.545, p = 0.46); (NPNGP: (F(1, 94) = 0.251, p = 0.62)].

To ensure the equity of prior academic achievement in each of the grading policy subgroups, we used the PR–GPA as an indicator. The ANOVA showed no significant differences between the grading policy subgroups in terms of the performance of the students in the prerequisite courses [F(2,315) = 1.144, p = 0.32)].

The ANOVA conducted to examine the differences between the three grading policies showed significant differences in final examination scores of the students [(F(2,315) = 4.421, p = 0.01]. Post Hoc comparisons using the Tukey–Kramer method [14] indicated that the average final examination scores with the TPOGP (M = 22.14, SD = 4.01) was significantly higher than the average scores with the OPOGP (M = 20.39, SD = 4.01), and the average examination scores with the NPNGP (M = 21.76, SD = 4.03) was not significantly different from the other two grading policies.

These results suggest that grading homework may not be a significant factor in improving the students' final examination performance since TPOGP vs. NPNGP, or OPOGP vs. NPNGP showed no statistically significant difference. However, the amount of homework assigned for collection had a significant impact on student examination performance since the comparison between TPOGP and OPOGP showed significant difference.

Figure 1 shows a histogram of student performance categorized as poor, below average, above average and good based on the number of correct answers to the 32 multiple-choice questions in the final examination. Note that the proportion of students who perform *good* or *above average* under OPOGP is considerably lower (43%) than the TPOGP (71%). While the NPNGP (with 60% of

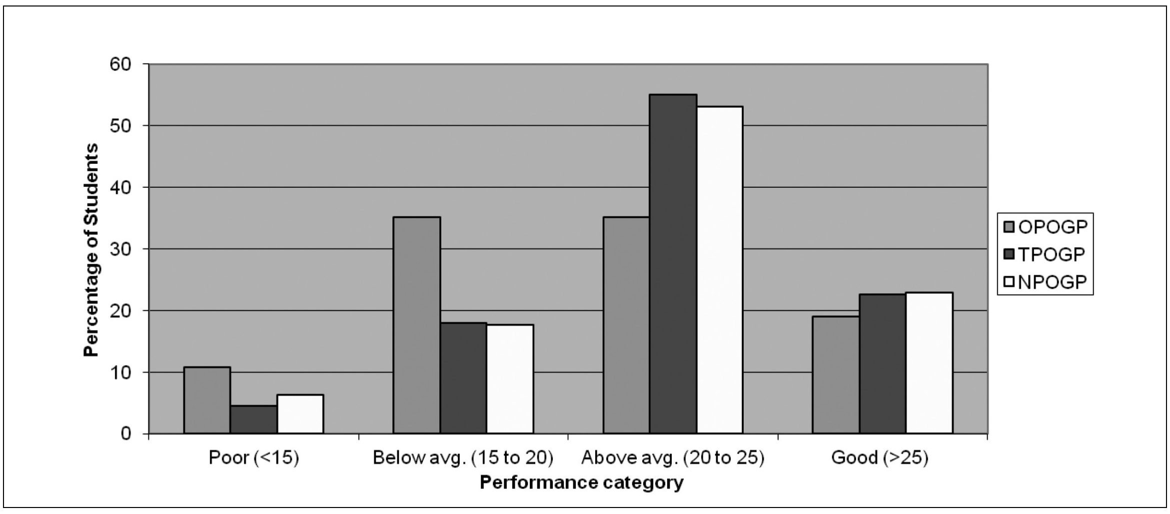

Fig. 1. Student performance histogram for different grading policies

the students performing as *good* or *above average*) was not significantly different from the other two grading policies, its performance was more comparable to the TPOGP than OPOGP.

To gain more insight, we continue analyzing the impact of grading policies with respect to the previously considered factors and subgroups.

3.5 *Analysis based on gender*

No statistically significant difference was observed between the male and female students' final examination performance in any of the homework grading policies. The corresponding ANOVA results, along with average final examination grades and standard deviations for each subgroup, are shown in Table 5.

3.6 *Analysis based on age*

The summary results for the analysis based on age subgroups are shown in Table 6. The only significant difference between the age groups was observed in the OPOGP where adult students performed worse than non-adult students under this policy. This result is interesting when it is viewed collectively with the fact that adult students overall performed worse than non-adult students in the final examination as previously pointed out in Section 3.3. This difference is only significant in the OPOGP, which clearly puts adult students at a disadvantage. In other grading policies, adult students' performance was much better and comparable to the performance of the non-adult students.

3.7 *Analysis based on transfer status*

Table 7 shows the summary analysis results based on the transfer status subgroups. There does not seem to be a statistically significant difference in the performance of the students in the various subgroups under any of the grading policies. However, it should be noted that CC students seemed always to perform the worst and the TPOGP seemed to help these students the most. The analysis of variance within each subgroup showed that while the FTIC and OT subgroups are not significantly

Table 5. Summary results based on gender

Grading policy	Gender	Number of students	Average final exam. grade	SD final exam. grade	ANOVA
OPOGP	Male	98	20.6	4.28	$F(1,109)<0.01$, $p = 0.99$
	Female	13	20.5	5.38	
TPOGP	Male	101	22.0	3.99	$F(1,109) = 1.257$, $p = 0.26$
	Female	10	23.5	4.22	
NPNGP	Male	85	21.6	4.06	$F(1,94) = 1.985$, $p = 0.16$
	Female	11	23.4	3.59	

Table 6. Summary results based on age

Grading policy	Age	Number of students	Average final exam. grade	SD final exam. grade	ANOVA
OPOGP	Adult	58	19.1	4.00	$F(1,109) = 15.645$, $p < 0.01$
	Non-adult	53	22.2	4.29	
TPOGP	Adult	59	21.8	3.88	$F(1,109) = 0.944$, $p = 0.33$
	Non-adult	52	22.5	4.16	
NPNGP	Adult	45	21.4	4.07	$F(1,94) = 0.741$, $p = 0.39$
	Non-adult	51	22.1	4.01	

Table 7. Summary results based on transfer status

Grading policy	Transfer status	Number of students	Average final exam. grade	SD final exam. grade	ANOVA
OPOGP	FTIC	50	20.9	4.05	$F(2,108) = 2.197$, $p = 0.12$
	CC	30	19.2	4.15	
	OT	31	21.4	4.95	
TPOGP	FTIC	57	22.3	4.15	$F(2,108) = 1.289$, $p = 0.28$
	CC	39	21.5	3.99	
	OT	15	23.4	3.42	
NPNGP	FTIC	41	21.8	3.56	$F(2,93) = 2.506$, $p = 0.09$
	CC	31	20.7	3.95	
	OT	24	23.1	4.61	

affected by the homework grading policy [$F(2,145) = 1.643$, $p = 0.20$; $F(2,67) = 1.485$, $p = 0.23$, respectively], the CC subgroup showed a somewhat significant difference [$F(2,97) = 2.852$, $p = 0.06$)]. Post hoc analysis using the Tukey–Kramer method showed no significant difference in the means, though it should be noted that the difference between TPOGP and OPOGP at 2.305 is very close to the minimum significant difference (MSD) at 2.327. This result is at least partially a consequence of the demographics of the students in our dataset. Approximately 81% of CC students are also adult students who were previously found to perform better under the TPOGP and the NPNGP.

3.8 Analysis based on prerequisite GPA

Summary results in Table 8 clearly demonstrate that students with PR–GPA $\geq$ 3.0 performed better under all grading policies. This is not surprising considering the previous results in Table 4 in Section 3.3 regarding this subgroup. Further analysis within each student group provided some interesting results. While the three different homework grading policies did not show a significant difference in the final examination grades of students with a PR–GPA $\geq$ 3.0 [$F(2,141) = 2.214$, $p = 0.12$], the effect of homework grading policy was a significant factor for students with PR–GPA < 3.0 [$F(2,171) = 4.004$, $p = 0.02$]. Post Hoc comparisons using the Tukey–Kramer method indicated that the average final examination scores with the TPOGP ($M = 20.9$, $SD = 3.85$) is significantly higher than the average scores with the OPOGP ($M = 18.9$, $SD = 3.93$). Figure 2 shows a histogram of student performance on the final examination for students with a course PR–GPA < 3.0. Note that the proportion of students whose performance was *good* and *above average* with TPOGP is visibly higher than the OPOGP. This difference is less prevalent in the case of the comparison between NPNGP and OPOGP, but is still discernable.

3.9 What type of student-learning is impacted?

We also collected data regarding student learning at the low and high levels of Bloom's taxonomy. The homework problems assigned for collection pertained mostly to low-level Bloom's taxonomy. Considering the significant differences in final examination performance regarding age, PR–GPA and homework grading policy, we further investigated these factors to determine if the difference in examination performance can be isolated to questions related to low or high levels of Bloom's taxonomy.

The results of the analysis based on age are shown

Table 8. Summary results based on PR–GPA

Grading policy	PR–GPA	Number of students	Average final exam. grade	SD final exam. grade	ANOVA
OPOGP	PR–GPA $\geq$ 3.0	52	22.4	3.93	$F(1,109) = 20.038, p < 0.01$
	PR–GPA < 3.0	59	18.9	4.20	
TPOGP	PR–GPA $\geq$ 3.0	45	24.0	3.64	$F(1,109) = 18.803, p < 0.01$
	PR–GPA < 3.0	66	20.9	3.85	
NPNGP	PR–GPA $\geq$ 3.0	47	23.3	3.52	$F(1,94) = 14.460, p < 0.01$
	PR–GPA < 3.0	49	20.3	4.00	

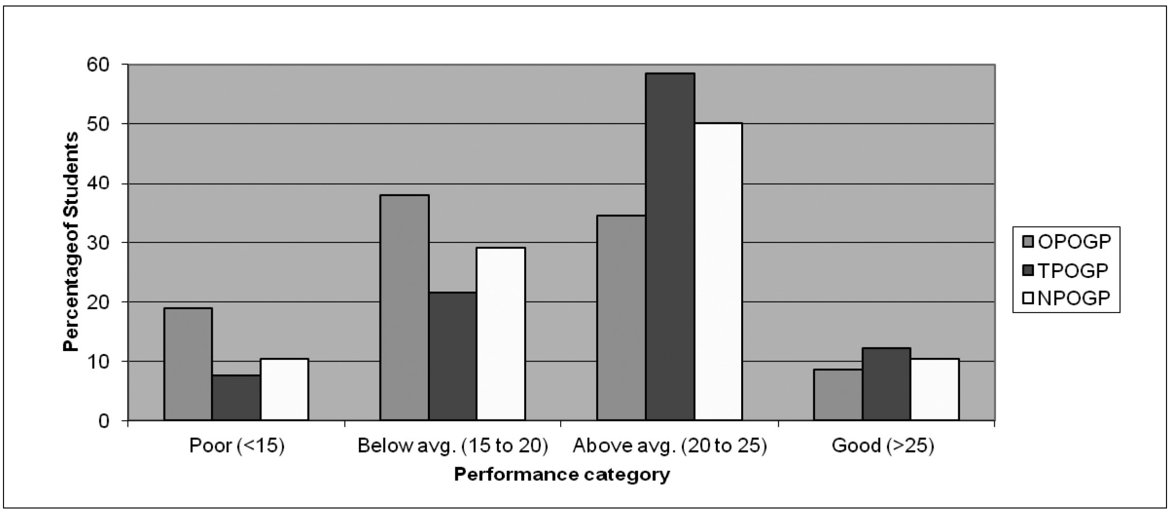

Fig. 2. Performance histogram of students with average PR–GPA < 3.0 for different grading policies.

Table 9. Summary results based on age

Learning level	Age	Number of students	Average final exam. grade	SD final exam. grade	ANOVA
Low	Non-adult	156	11.4	2.37	F(1,316) = 7.396, $p < 0.01$
	Adult	162	10.7	2.34	
High	Non-adult	156	10.9	2.25	F(1,316) = 9.903, $p < 0.01$
	Adult	162	10.0	2.53	

Table 10. Summary results based on PR–GPA

Learning level	PR–GPA	Number of students	Average final exam. grade	SD final exam. grade	ANOVA
Low	GPA $\geq$ 3.0	144	11.8	2.27	F(1,316) = 29.380, $p < 0.01$
	GPA < 3.0	174	10.4	2.29	
High	GPA $\geq$ 3.0	144	11.4	2.14	F(1,316) = 44.749, $p < 0.01$
	GPA < 3.0	174	9.6	2.39	

in Table 9 and indicates that the difference in performance was in both the low and the high level questions. The same conclusion can be drawn from the analysis based on the PR–GPA factor. The results of this analysis are shown in Table 10.

The analysis regarding homework grading policy produced more interesting results than the age and PR–GPA factors. The ANOVA conducted shows no significant difference between the three grading policies regarding the number of correct answers to questions at the high level of Bloom's taxonomy [F(2,315) = 1.823, $p = 0.16$]. On the other hand, there is a significant difference between these grading policies when the correct answers to questions at the low level of Bloom's taxonomy are considered [F(2,315) = 5.615, $p < 0.01$]. Post hoc comparisons using the Tukey–Kramer method indicated once again that the average final examination scores with the TPOGP (M = 11.5, SD = 2.28) is significantly higher than the average scores with the OPOGP (M = 10.5, SD = 2.44). No significant differences in average examination scores were detected between NPNGP and the other two grading policies. Figure 3 shows student performance on low level Bloom's taxonomy questions for different grading policies. Note again the significantly higher proportion of students whose performance was above average or good for the TPOGP and the NPNGP case. Clearly the OPOGP was not as favorable as the other two grading policies as far as correctly answering the questions at the low level Bloom's taxonomy are concerned.

3.10 *Summary discussion of results*

Similar to the previous work [5–6] in this area, our results indicate that collecting homework for a

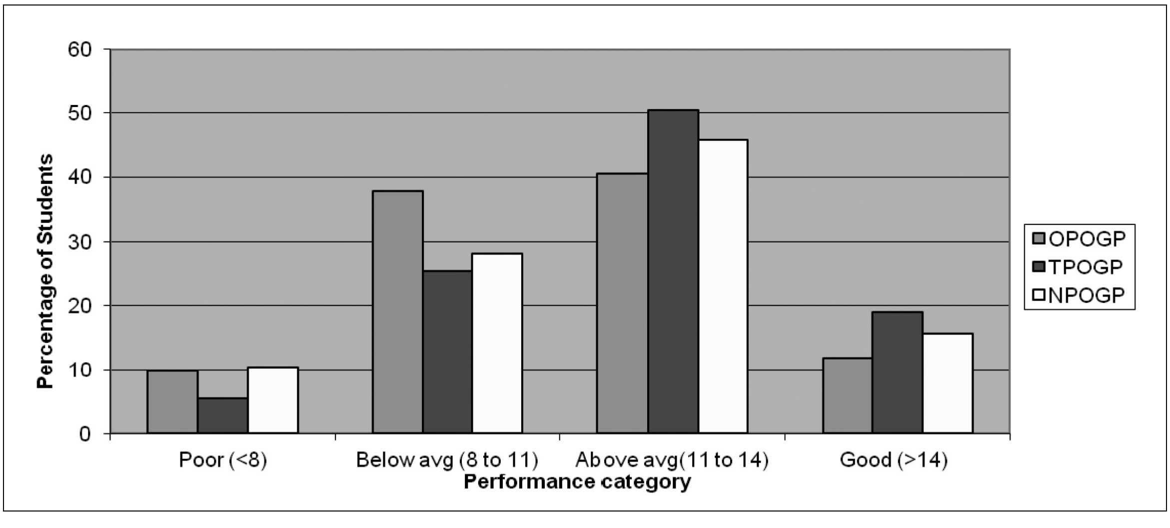

Fig. 3. Performance histogram of student performance on low level Bloom's taxonomy questions for different grading policies.

grade does not affect student performance. Nowhere in our analysis were we able to identify a subgroup or subgroups that performed differently under the NPNGP and the TPOGP or OPOGP. However, we did notice that the grading policy OPOGP, where one homework problem is assigned for collection and graded, performs significantly worse than the grading policy TPOGP, where three problems were assigned for collection and one problem was graded. We believe that this is simply because one-problem homework makes it extremely easy for students to copy and such practices are known to greatly degrade student performance [18–19].

The most successful grading policy is where three homework problems are assigned for collection and only one is graded, especially for particular subgroups such as the students with a PR–GPA < 3.0 and adult students. In almost all cases, this policy had better average performance than the no graded homework policy, albeit not statistically significant.

Based on the overall analysis of the data, we are inclined to believe that student performance is attributed to the amount of practice students are getting. In the NPNGP, the students practiced during class time through the use of personal response systems for reviewing pre-requisite knowledge and checking end-of-topic knowledge. In the TPOGP, one out of three problems was graded but the additional homework problems students had to solve provided the extra practice. Students under the OPOGP either did not get sufficient practice or failed to devote the necessary time and effort to the problem, and therefore performed consistently worse than the students subjected to other grading policies. This belief is further supported by the finding that students with a PR–GPA ≥ 3.0 were not affected by the grading policy at all as these students are believed to be sufficiently motivated to get adequate practice without the graded homework incentive.

Another interesting observation from the data is that there was no subgroup that was more or less inclined to turn in homework. Students in all subgroups submitted on average 88–92% of the homework, which is fairly high. The high submission rate may be attributed to the junior level standing of the students involved, most of whom have mastered effective study habits. The latter may also be one of the reasons why the no graded homework policy performed comparably to graded homework policies as students knew what and how much to study to master the material.

In terms of student learning, we detected no significant difference between the grading policies for questions in the high levels of Bloom's taxonomy. These are the types of questions that were *not covered* in either the homework problems assigned for collection or the in-class exercises used in the NPNGP. However, questions in the low level Bloom's taxonomy (similar to the homework assigned for collection and in-class exercises) were answered better by the students in the TPOGP. In terms of the comparisons within age and PR–GPA subgroups, significant performance difference was at both the high and low levels of the Bloom's taxonomy.

The results of this summary discussion must be considered in conjunction with the limitations of the study. The students considered in the study are junior level Mechanical Engineering students who are required to have a 2.5 GPA in mathematics and natural science course requirements to be accepted to the program. As mentioned before, these students are more likely to have some level of established effective study habits compared with entry level freshman and may perform equally well when the structure provided by weekly homework is absent. Furthermore, the course in which the study was conducted is a medium size course (40–70 students), which allows some degree of in-class hands-on exercises and a higher degree of in-class discussion compared with very large courses with 200–300 students. The generalization of these results requires a more extensive study involving students from different departments, courses at different levels, and different class sizes.

4. Conclusions

In this paper, we reported on an in-depth study to understand the effects homework grading policies on student performance measured by a multiple-choice final examination. The study is based on data from more than 300 students over a period of three years in six different semesters taught by the same instructor where the only difference over the semesters was the homework grading policy. We maximized the reliability of the data by gathering it from official student records and using no self-reported information. The primary performance measure used in the study was a multiple-choice examination, which effectively eliminates any possible grading bias.

We found no conclusive evidence that grading homework improves student performance within the specific cohort of students involved in the study. Our results also indicate that subjecting students to practice problems either in class or as homework significantly improves student performance especially for some student subgroups. However, even if there does not seem to be convincing evidence that graded homework improves student performance, faculty use graded homework as an early indicator of how the overall class is doing and the students' level of understanding. Understandably, faculty may be reluctant to eliminate graded homework. However, if resources are scarce, and faculty do decide to eliminate graded homework, results from this study show that we are not doing a disservice to our students as long as sufficient opportunities for practicing new concepts are available and expected.

Acknowledgments—This material is based upon work partially supported by the National Science Foundation under Grant No 0717624, and the Research for Undergraduates Program in the USF College of Engineering. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. H. Cooper, *Homework*, Longman, White Plains, NY, 1989.
2. H. Cooper, J. C. Robinson and E. A. Patall, Does homework improve academic achievement? A synthesis of research, 1987–2003, *Review of Educational Research*, **76**(1), 2006, pp. 1–62.
3. C. Henderson, K. Heller, P. Heller, V. Kuo and E. Yerushalmi, Students learning problem solving in Introductory Physics: Forming an initial hypothesis of instructor's beliefs, *AAPT Physics Education Research Conference*, Boise, ID, 2002.
4. L. A. Theodore, T. L. Hughes, R. J. DioGuardi, D. Aloiso, M. Carlo and D. Eccles, A class-wide intervention for improving homework performance, *Journal of Educational & Psychological Consultation*, **19**(4), 2009, pp. 275–299.
5. G. Weems, The impact of homework collection on performance in intermediate algebra, *Research and Teaching in Developmental Education*, **15** (1), 1998, pp. 21–25.
6. H. Trussell and E. Dietz, A study of the effect of graded homework in preparatory math course for electrical engineers, *Journal of Engineering Education*, **92**(2), 2003, pp. 141–146.
7. M. H. Peters, R. B. Kethley, K.E. Bullington and K. Kim, The impact of homework on student performance in an introductory production and operations course, *Journal of Business Education*, **1**, 2000, http://www.abe.sju.edu/proceeding.html, Accessed Jan. 24, 2011.
8. B. Gutarts and F. Nain, Does mandatory homework have a positive effect on student achievement for college students studying calculus? *Mathematics and Computer Education*, **44** (3), 2010, pp. 232–44.
9. S. Bonham, R. Beichner and D. Deardorff, Online homework: does it make a difference? *The Physics Teacher,* **39**(5), 2001, pp. 293–296.
10. K. J. Burch and Y.J. Kuo, Traditional vs. online homework in college algebra, *Mathematics and Computer Education*, **44**(1), 2010, pp. 53–63.
11. D. Doorn, S. Janssen and M.O'Brien, Student attitudes and approaches to online homework, *International Journal for the Scholarship of Teaching and Learning*, **4**(1), 2010, pp. 1–20.
12. A. Chickering and Z. Gamson, Seven principles for good practice in undergraduate education, *AAHE Bulletin*, **39**(March)**,** 1987, pp. 3–7.
13. B. Bloom, *Taxonomy of educational objectives, Handbook I: Cognitive Domain*, David McKay Co., New York, 1956.
14. D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons Inc., New York, 2001.
15. L. R. Gay and G. E. Mills, *Educational Research: Competencies for Analysis and Applications*, 8^{th} edn, Prentice Hall, 2005.
16. P. Wankat, The role of homework, *Proceedings of the American Society for Engineering Education Annual Conference & Exposition*, 2001, Albuquerque, NM.
17. A. Fernandez, C. Saviz and J. Burmeister, Homework as an outcome assessment: relationship between homework and test performance, *Proceedings of the American Society for Engineering Education Annual Conference & Exposition*, 2006, Chicago, IL.
18. D. Palazzo, Y. Lee, R. Warnakulasooriya and D. Pritchard, Patterns, correlates and reduction of homework copying, *Physical Review Special Topics Physics Education Research,* **6**(1), 2010, pp. 010104-1–010104-11.
19. L. Sanders, Homework makes the grade, *Science News*, 2010, http://www.sciencenews.org/view/generic/id/57656/title/Homework_makes_the_grade, Accessed Jan. 24, 2011.

Ali Yalcin is an Associate Professor of Industrial and Management Systems Engineering Department at the University of South Florida, and an Associate Faculty member of the Center for Urban Transportation Research. His research interests include modeling, analysis and control of discrete event systems, production planning and control, industrial information systems, data analysis and knowledge discovery, and engineering education research. He has taught courses in the areas of systems modeling and analysis, information systems design, production planning, facilities design, linear systems and systems simulation. He also co-authored the 2006 Joint Publishers Book-of-the-Year textbook, *Design of Industrial Information Systems*, Elsevier. He is a member of the American Society of Engineering Education (ASEE). He received his B.S. in 1995, M.S. in 1998, and Ph.D. in 2000 in Industrial and Systems Engineering from Rutgers University, NJ.

Autar Kaw is a Professor of Mechanical Engineering and Jerome Krivanek Distinguished Teacher at the University of South Florida. His main scholarly interests are in engineering education, bridge design, composite materials, and the state and future of higher education. Autar has written several textbooks on subjects such as composite materials, numerical methods, and computer programming. Since 2002, under his leadership, he and his colleagues from all over the US have developed, implemented, refined and assessed online resources for open courseware in Numerical Methods (http://numericalmethods.eng.usf.edu). In 2010, there were more than a million page views of http://numericalmethods.eng.usf.edu, and more than half-a-million views of the audiovisual lectures on YouTube at http://youtube.com/numericalmethodsguy. His work has appeared in the St. Petersburg Times, Tampa Tribune, Chance, Oracle, and his work has been covered in Campus Technology, Florida Trend Magazine, and ASEE Prism. Autar is a Fellow of the American Society of Mechanical Engineers (ASME) and a member of the American Society of Engineering Education (ASEE). He received his B.E. Honors degree in Mechanical Engineering from Birla Institute of Technology and Science (BITS), India in 1981, and his Ph.D. in 1987 and M.S. in 1984, both in Engineering Mechanics from Clemson University, SC. Professor Kaw received the ASEE National Teaching Medal in 2011 and the Florida U.S. Professor of the Year Award from the Council for Advancement and Support of Education (CASE) and Carnegie Foundation for Advancement of Teaching (CFAT) in 2004.