SELF-GRADED HOMEWORK, GENDER, AND STUDENT COMPREHENSION: WHAT'S THE CONNECTION?

Submitted to the 2012 DSI Annual
Instructional Innovation Award Competition

April, 2012

SELF-GRADED HOMEWORK, GENDER, AND STUDENT COMPREHENSION: WHAT'S THE CONNECTION?

ABSTRACT/INNOVATION SUMMARY

Allowing students to grade their own homework promises many advantages. Benefits to *students* include (1) the availability of immediate feedback, (2) the potential for enhanced student engagement, (3) a greater sense of ownership in their studies, (4) a friendlier, more productive, and cooperative classroom, environment, and (5) a shared sense of ownership for the learning process. Benefits to *instructors* include (1) a surrogate for taking attendance, (2) the ability to "count" work that they might otherwise not require, or only assign on an optional basis, (3) the savings in time that accrues when instructors simply record self-reported scores instead of grading the homework assignments themselves, (4) the ability to assign more homework, and (5) the ability to assign homework in mass-lecture venues.

Allowing students to grade their own homework also raises concerns. Two key questions are whether students can perform such grading tasks honestly and whether they can grade them accurately. Another question is whether such work varies by gender.

Most empirical studies of student self-assessment have been conducted in the social sciences. Similar studies in engineering, computer science, or business administration are notable for their absence. To address this deficiency and answer these questions, the author allowed the students in his introductory programming classes to grade their own homework—at least 20 assignments in each of seven such classes. He then analyzed the homework scores, final examination scores, and the effects of gender of the 266 students in these classes. The statistical results were favorable to the hypothesis that students graded themselves fairly, and he also found no gender bias in this equity. He did, however, find marked gender differences in the final exam scores.

Not every class necessarily lends itself to student self-grading, but many business courses that require step-by-step, sequenced logic to arrive at a homework answer—for example, courses in computer programming, finance, logistics, supply chain management, or accounting—probably do. The results of this study suggest that students can be trusted to grade their own homework, and perhaps with the assistance of grading rubrics or other evaluation aids, are able and willing to do so. These findings therefore promise much to those instructors seeking ways to deliver quality education in large business classes as well as to those budget-strapped academic institutions seeking ways to "do more with less."

INTRODUCTION AND RELEVANT LITERATURE

The term "collaborative learning" (CL) describes a wide range of activities in which students assume responsibility for some of the educational activities in their courses. One example of collaborative learning is participation in teams to accomplish such tasks as homework assignments, case analyses, project presentations, or even examinations (Cortright, et al., 2003; Mumford, 2010; Michaelson, et. al, 2004). Another example is when students contribute in some way to class learning, as for example, when they lead discussions about specific course subjects, participate in peer-evaluation tasks, or suggest examination questions for class tests (Sadler and Good, 2006).

A growing body of both anecdotal and empirical evidence suggests that collaborative learning applications enjoy many advantages. For example, Koppenhaver (2006) argues that CL tasks requiring teamwork increase the opportunities for cooperative work, help students learn how to reach collective compromises, improve interpersonal skills, and facilitate group problem solving. Similarly, Vander Schee (2011) suggests that allowing students to select the weights used to determine their final course grades positively influences their commitment to their courses, their sense of control of their classes, and even course performance. Finally, Iqbal, Kousar, and Rahman (2011) note that collaborative learning exercises may be effective strategies in distance learning environments, in which onsite, face-to-face interactions are limited.

A particularly-interesting dimension of collaborative learning to the authors is a system that allows students to grade their own homework. This approach would seem to challenge instructors in courses requiring integrative analyses, theory syntheses, or interpretive skills. But, empirically, this conclusion appears to change in courses where the major homework objective is to familiarize students with problem-solving techniques and where students are provided with a grading rubric with which to evaluate their work—especially the amount of points to deduct for common errors (Bound, 1989). Similarly, such procedures may become more feasible where homework problems have strict, right answers such as in accounting or the STEM disciplines (science, technology, engineering, and mathematics).

At face value, "student self-grading" promises several advantages to both students and instructors. One benefit to *students* is the potential to enhance their engagement and perhaps commitment to the learning goals of a course. If student homework grading takes place on the day it is due, such self-grading also offers immediate feedback—a factor believed to positively influence learning and retention (Edwards, 2007). Student self-grading also provides an opportunity for students to deepen their understanding about a subject—for example, to better understand why their answer is wrong, or why an alternate answer is superior (Sadler and Good, 2006). Finally, empirical evidence suggests that self-grading improves class attendance, makes the classroom experience a friendlier, more productive, and cooperative environment, and provides a shared sense of ownership for the learning process (Strong, et. al., 2004).

One potential benefit of student self-grading to *instructors* is that submitted work acts as a surrogate for taking attendance. Then too, self-graded homework allows instructors to "count" work that they might otherwise not require, or only assign on an optional basis—a policy that Chickering and Gamson (1987) list as one of the seven principles of good practice in undergraduate education. A third advantage is the time that instructors save when they merely record self-reported scores instead of grading the homework assignments themselves (Sadler and Good, 2006). Yet a further advantage is an instructor's ability to assign more homework—a particularly useful advantage to the teachers of mass-lecture classes such as large introductory courses. Finally, self-graded homework policies have the potential to increase student engagement in coursework, if only because self-grading transforms students from passive submitters of work to active evaluators of such work.

Just because instructors *can* allow students to grade their own homework does not mean that instructors *should* install such a policy. After all, not all university classes necessarily lend themselves to graded homework, some students balk at exercises that they accurately or wrongly consider "busy work" or "not my job," and some instructors feel that their in-class tests adequately motivate students to do whatever homework is necessary to master course materials. How important is homework in college learning anyway?

Twenty-five years ago, the general consensus was that student learning gains were directly related to the amount of time students spent on learning—time that increases with assigned and graded homework (see, for example, Sorcinelli, 1991). More recent empirical tests have not always confirmed this thinking. For example, studies by Cooper (1989), Vruwinkle and Otto (1987), Peters, Kethley, and Bulligan (2002), and Geide-Stevenson (2009), all found that required, graded homework in college courses did not increase achievement metrics when compared to those of control groups in classes not requiring such work. In contrast, studies by both Weems (1998) and Grove and Wasserman (2006) found the opposite result—i.e., that graded homework *did* appear to result in measurable learning gains.

Several additional factors also negatively influence the advisability of self-grading. One of the most onerous concerns is the amount of expertise required in the grading process itself. How can students adequately grade their own work in those courses covering unfamiliar material? Several authors suggest that they cannot—see, for example, Andrade and Du (2007) and Kirby and Downs (2007). But a growing body of empirical evidence suggests the opposite. For example, a study by Boud and Falchikov (1989) found that most student marks agreed with those of their teachers. Similarly, a study by Stefani (1994) found that student self-assessed grades were similar to those of their tutors. Finally, Leach (2012) found no statistical difference between the mean student (self-assessed) grade of 5.57 and the mean teacher grade of 5.58 ($p<.01$) for the homework materials of 120 students in her adult education classes.

Lastly, there is the matter of "honesty" in student self-grading. Even if students are capable of evaluating their work objectively does not guarantee that they will do so. Moreover, if instructors include self-graded homework as a component of their final course grades, there is an obvious incentive for students to be generous in grading their own works, or in simpler terms, to cheat on them (Andrade and Du, 2007; Kirby and Downs, 2007). This concern is why some experts believe that formal self-assessments are not more widely used in higher education (Kirby and Downs, 2007; Thompson, et. al., 2005).

A growing body of empirical investigations appears to confirm this belief. For example, Sadler and Good (2006) compared student homework evaluations with teacher grades for the same work in four of their general-science classes and found that "lower-performing students tended to inflate their own low scores." Similarly, a study by Leach (2012) of 472 students made these same observations for lower-achieving students, but also found that higher-achieving students tended to underrate themselves. Yet a third study by Strong, Davis, and Hawks (2004) of 480 students in their history classes also found statistical evidence of grade inflation: 57 percent of self-assessments resulted in "A" grades, compared to 31 percent of "A" grades when only teachers assigned grades. Finally, using an anonymous survey, Edwards (2007) found that as much as 20 percent of the students in a social statistics class reported that they "had actually seen other people cheating" at least some of the time. Interestingly, as much as 95 percent of the students in that same study claimed that they, themselves, had "never cheated."

INNOVATION

There are many dimensions within the broad topic of student self-assessment, including several that have yet to be investigated. One simple matter is that most empirical studies of student self-assessment have been conducted in the social sciences. Similar studies in engineering, computer science, or business administration are notable for their absence. Thus, one rationale for investigating self-graded homework in colleges of business is to examine its potential in a new venue. Then, too, it is currently not known whether the effectiveness or trustworthiness of self-assessments varies by gender. These are the questions addressed by the current study.

To examine these dimensions empirically, the author allowed the students in several of his introductory programming classes in the college of business of a public, western, land-grant university to grade their own homework. The homework due, which was due virtually every day in class, required students to create as many as five small programming applications in Visual Basic from the end-of-chapter exercises in Snyder (2011). The number of such assignment sets necessarily varied from semester to semester, but was never less than 20 assignments.

IMPLEMENTATION

The experiment ran for 7 consecutive semesters. At the start of each class period for any given semester, students were shown a suggested solution to each of the individual problems assigned for that day, as well as a table indicating the maximum number of points to award for each exercise. Students were also allowed to award themselves partial credit for work if they felt they had earned it, and some students awarded themselves credit in as little as tenths of a point.

In all these classes, the instructor stressed the importance of doing all the homework because "learning how to use a procedural programming language" was a primary learning objective of the class. For this reason, and to encourage students to do it, the unchallenged, self-graded homework counted for either 20 or 25 percent of a student's final grade in each semester. The Observations section of this paper contains further comments on this policy.

Table 1 provides some statistics for each of the seven classes in the sample. The total number of students was 266, of which 80 students were female and 186 were male. The average of the total number of points for all homework points, per class, was slightly more than 255 points. Again, some problems in these homework sets were worth as little as one or two points, while others were worth as much as seven or eight points.

Table 1: Selected Class Statistics

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Number of Students	36	31	39	39	40	40	41
Number of Females	8	5	14	13	13	13	14
Number of Males	28	26	25	26	27	27	27
Maximum homework points	286	229	253.5	256	271.5	256	234.5
Homework as percent of final grade:	20%	20%	20%	25%	25%	25%	25%

To examine the viability of student self-graded homework scores, the author used a linear regression model of the form:

$$\text{Final} = \alpha + \beta\text{Gender} + \delta\text{Homework} + \varepsilon$$

Where:

Final = final examination score on a 50-question, multiple choice test (scaled to a percentage);
α = a constant;
Gender = male or female, coded as female = 1 and male = 0;
Homework = total points earned on the self-graded homework (Table 1), scaled to a percentage; and
ε = an error term.

The final examinations in each class were 50-question, multiple-choice tests mostly drawn from the test bank of the (Schneider, 2012) class textbook and varied little from one class to another in content (the instructor collected test booklets at the end of each final exam). In the author's opinion, these were straightforward questions that tested student knowledge on the programming constructs practiced in the homework.

MEASURING EFFECTIVENESS

If students had accurately and honestly graded themselves, we would expect to see a close statistical relationship between their homework and the scores on their final exam, which was based on such work. Conversely, the absence of a statistical relationship between the two leads to the conclusion that self-grading may not be advisable. Finally, the gender variable adjusts for potential differences in test performance attributable to that variable.

We used a Minitab statistical package to estimate all statistical results for our study. The estimated coefficients for our linear regression were:

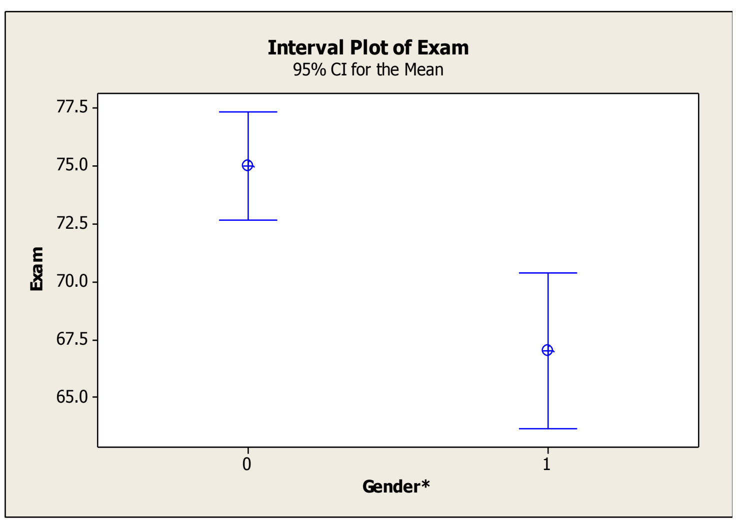
Final Exam = .57 - .071Gender + .265 Homework; R^2 = .146, F = 23.67.

All parameter estimates were statistically significant at p = .000 level of significance. The value of the Homework coefficient was positive as desired, indicating a direct relationship between student self-graded homework scores and student final exam test scores. The value of .265 means that, for every one percent increase in homework scores, students would obtain about a quarter-of-a-percent increase in their test score. If students cheated on their homework, we would *not* expect to see a statistically significant coefficient here—homework scores would have less predictive capability. This provides at least some evidence for a lack of misrepresentation on homework scores.

The estimated gender coefficient of -.071 means that, on average, females scored about seven percent lower on their final examinations than did males. This result may be disappointing to those wishing to find gender neutrality on tests, and contrasts with Picou (2011) who also tested "gender" as a possible predictor of test performance but found no statistical significance for this variable.

Possible reasons for the observed, statistically-significant differential between males and females in our regression equation are problematic. For the sample at hand, in fact, these gender differences were striking. Figure 1 depicts 95% confidence intervals for the final exam scores of the male and female students in this study. Neither the means nor the confidence intervals for these two sets of individuals overlap, providing striking visual evidence for what our regression equation estimates have already stated—men and women were *not* equal in their test performances.

Figure 1. Ninety-Five percent confidence intervals for male and female student performance on the final examination.



One, older rationale for this differential is simply that "males are better than females" in the STEM sciences—an argument with substantial empirical counter-evidence today (Begley, 2012). Another possible reason for our observed gender differentials is the possibility of sampling error—i.e., the chance that we simply drew a biased sample. However, the likelihood of this also seems small, given that both male and female students self-selected into the course and our sample spanned 7 semesters. This leaves at least one additional possibility: a gender bias in multiple-choice question test formats. Here, a stream of research studies has found similar results—i.e., suggest that women do slightly poorer on MC tests than do men (see, for example, Simkin and Kuechler, 2010)—perhaps the most likely explanation observed here.

The R^2 value of .146 was disappointingly small, suggesting that our model explained less than 15% of the variance in the dependent variable. While such results are not uncommon in behavioral models of this type, its low magnitude suggests that much of what explains student achievement on final exams is attributable to variables that were not included in our model—for example, age, native intelligence, commitment or interest to the subject matter of the course, absenteeism, or even guessing ability on multiple-choice tests (Picou, 2011).

The authors also performed two additional statistical analyses: (1) a set of linear regressions for each class individually, and (2) separate regressions for males and females. The least-squares results for each class taken separately mostly mirrored those presented here for the combined set of data—i.e., the gender coefficient was consistently negative and the homework coefficient was consistently positive. R^2 values varied between 0 and 30 percent.

The primary reason for running separate regressions for male and female students was to see if self-grading varied by gender. Table 2 provides results for this analysis. Again, the homework variable was positive and statistically significant at $p = .000$. Values for R^2 were again small, suggesting that both equations had limited predictive power, but also inferring that neither gender's self-graded homework was much better than the other for that task.

Table 2: Linear Regression Results for Males and Females

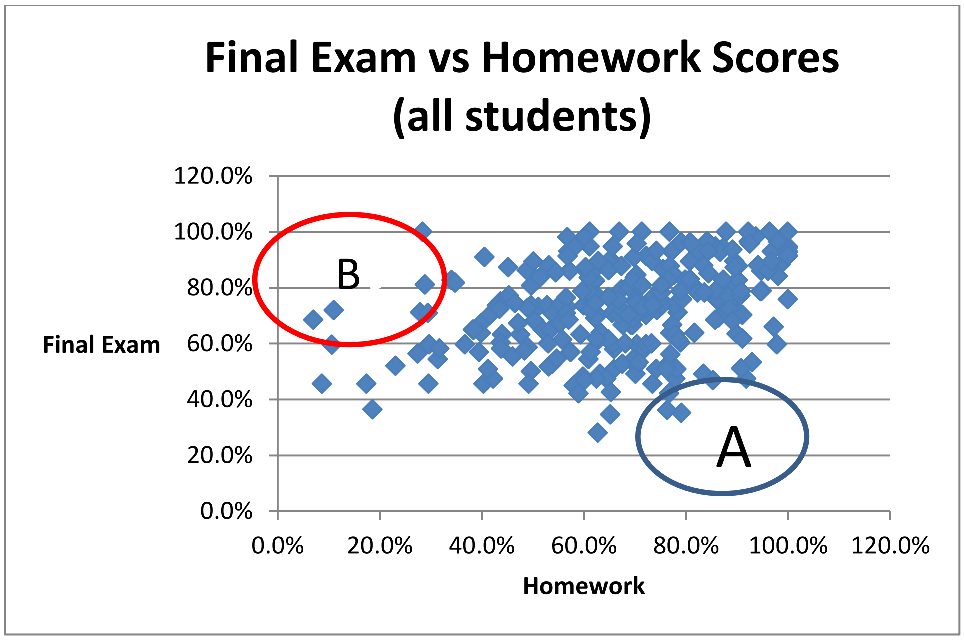
	Constant (P-value)	Homework (P-value)	R^2
Females	48.60 (.000)	.28 (.000)	.15
Males	57.49 (.000)	.26 (.000)	.09

Looking for Outliers

In the opinion of the authors, the best evidence for student largess on their homework lies not in the predictive properties of the model discussed above, but in the potential outliers in the data. In particular, if students had misrepresented their homework scores in any systemic way, we would expect to find some students with high homework scores and low exam scores—i.e., we would expect to see outliers in the area marked "A" in the scatter diagram of Figure 2. The authors did find outliers, but not where they expected to see them. Instead, most of the outliers fell in area B—i.e., an area characterized by low homework scores but high final-exam scores. How could this be?

The author was familiar with most of the students in his classes, and in particular, those students who fell in area B of Figure 2. For the most part, these were smart students who did well on tests but who were disinclined to do much homework. For example, several of the males in this group were computer science, rather than business, majors and were taking the course as a casual elective. In contrast, some of the females in this group were working "supermoms" with little free time and higher priorities elsewhere. Collectively, therefore, most of the individuals in area B simply did not have strong incentives to do much homework.

Figure 2. A scatter diagram of the sample data



TRANSFERABILITY AND IMPLICATIONS FOR EDUCATORS

Several matters concerning this study are worthy of further comment. Four of them are (1) the questionable policy of counting self-graded homework as a sizable percentage of a student's final grade, (2) the statistical viability of data that have the potential to fall within narrow ranges, (3) the appropriateness of student self-grading in programming classes, and (4) the mechanics of in-class grading. The paragraphs below address each of these matters in turn.

In the present study, the author not only allowed his students to grade their own homework, but *counted* self-graded homework in the computation as component of the final course grade. This might seem inadvisable, but the author suggests that this policy may not be as risky as might first appear. The practice first began simply because his students complained that the programming assignments were a lot of work and argued that they "deserved credit for it." But students quickly realized on their own that any large disconnect between their homework scores and either the midterm of final exam scores were grounds for a serious, one-on-one discussion in the professor's office. Probably for this reason, no such discussions were ever necessary, although this is an inference on the instructor's part.

It is also worth noting that most of the students in this class were juniors and that the median age for students at this university is "26." In addition, at the start of each semester, the instructor announced that he was placing students on the honor system regarding their self-grading, but that their grades would not be challenged. In practice, it was the author's observation that many of his students were harder on themselves than he would have been.

With regard to the "narrow-range problem," a potential concern in statistical analyses is a possible lack of variability in the data for independent variables in linear regression work. This is a problem because low variability in the sample data makes it difficult to estimate regression parameters with any degree of precision. At least for the sample used here, this fear was (surprisingly) unrealized. In class 1, for example, homework scores ranged from a low of "53" to a high of "229." Each of the other classes displayed similar, wide ranges. Such variability also speaks to the integrity of the grading process to the extent that we would not expect to see such dispersions if student "generosity" in self-grading was widespread. The scatter diagram in Figure 2, which displays the homework scores for the entire sample, really tells the tale: "variability" was simply not a statistical issue here, despite the potential—and perhaps a priori fear—that it would be a problem.

With regard to the question of the appropriateness of a self-grading policy, the authors realize that this is ultimately an issue best left to the individual instructor. Certainly, there are some classes such as seminars or "current issues" classes in which "graded homework" may not make sense. In like manner, an instructor who announces that his or her in-class examinations are based on problems similar to those in the suggested (but ungraded) homework may find that formally *requiring* such homework doesn't make much difference in test performance—students will have taken the hint to heart and do the homework whether it's graded or not. Additional factors that may influence the effects of homework policies may include the subject domain of the course, the instructor, the general course policies, the metric used to measure learning gains, or any of a long list of additional factors whose effects have not been measured and therefore are not fully understood. These are additional factors for further research.

In the subject domain of the present study, there are usually several methods for achieving a given programming objective, and perhaps the most common question from students during the "grading portion" of a given class period in the study classes was the appropriateness of an alternate solution. In the vast majority of cases, the author found that students had used novel, and often amazingly creative, approaches to solve homework problems. In the author's opinion, and in aggregate, the ensuing class discussions about such alternate tactics was one of the most educational portions of the course—especially for him! Self-grading in this sense created additional learning opportunities.

Finally, we make several brief qualitative observations about the mechanics of the self-grading process itself. First, we note that in-class grading takes valuable class time—an opportunity cost that some instructors may not wish to incur. Secondly, we confirm what others have already noted about the psychological impact of self-grading. Commentary on end-of-semester class evaluations from students suggests that most students liked the process, appreciated the "vote of confidence" in their integrity, and it would appear, lived up to instructor expectations of honest grading. Lastly, we observe that, when the grading happens at the beginning of class, the process encourages students to come to class prepared and on time.

CAVEATS AND DIRECTIONS FOR FUTURE RESEARCH

Unlike several of the prior studies conducted to date, there was little evidence in the sample data of the current study for student fudging on homework. Nonetheless, the absence of statistical evidence for grade misrepresentation does not necessarily mean that none occurred.

The underlying coursework for this study were classes in Visual Basic—i.e., an entry-level course in computer programming—and students had ample opportunity to obtain outside help from a number of sources, including the teaching assist for the course, parents who happened to be in the IT profession, students who had taken the course in prior years, and of course, each other. Beyond this, the author's observation is that each class probably had some students who resorted to trial and error to complete their assignments, meaning that they understood little but (after a sufficiently high number of attempts) *somehow* got their programs to work. In all such cases, the end results were completed homework assignments deserving of high grades for *achievement*, but not *comprehension*. Given this consideration, it is almost surprising that the current study found any relationship between "self-assessed homework scores" and "exam grades." This is an important concern for future research, inasmuch as most studies of student assessments have measured the former metric ("achievement"), not the latter ("comprehension"). This is one reason why future researchers might prefer to use independent tests of student *comprehension*, measured by objective testing, rather than "parallel grading by others," when assessing the effectiveness of the self-grading process.

It is also important to note that, in the current study, students always graded themselves with a grading rubric and, sometimes, also had comments from the instructor about why a given answer did or did not deserve full credit. The authors realize that not every class lends itself to such grading assistance, or that every instructor might feel that much class time should be devoted to it. Either way, student self-grading becomes a completely different animal without it.

SUMMARY AND CONCLUSIONS

One dimension of "collaborative learning" is a policy that allows students to grade their own homework. But can students do this accurately and honestly? A growing body of conflicting evidence is one reason why we need yet further studies to investigate these questions—especially in disciplines where the advantages of self-graded homework are promising but the evidence is either weak or non-existent. A related issue is whether self-grading varies by gender.

The present study used the aggregate homework and final examination scores from 247 students to address these questions. We found a positive, statistically-significant relationship between these two variables, suggesting that, if students misrepresented their homework scores in any substantive way, it was not statistically identifiable in our analyses. We also included "gender" in our analyses, which (while small) was also statistically significant in predicting performance on an objective final examination. Because of its small size, we are not inclined to make further inferences about the honesty of the underlying population.

Finally, the authors expected to find outliers in the sample data—in particular, students with high homework scores but low exam scores. The presence of such outliers would have been the clearest signal of “generosity” in student grading, and *prima facie* evidence against the practice. In this study, we found outliers, but in exactly the opposite place we expected them—i.e., students with low homework scores and substantially higher exam scores.

REFERENCES

Andrade, H. and Y. Du (2007). Student Responses to Criteria-Referenced Self-Assessment. *Assessment & Evaluation in Higher Education,* vol. 32, no. 2, pp. 159-181.

Begley, Sharon (2012). “Anything Boys Can Do…” *Scientific American* vol. 306, No. 1 (January), p. 25.

Berry**,** John and Melvin A. Nyman, (2002), “Small-Group Assessment Methods in Mathematics” *International Journal of Mathematical Education in Science and Technology* vol. 33, no. 5, pp. 641-9.

Boud, D. (1989). “The Role of Self-Assessment in Student Grading” *Assessment and Evaluation in Higher Education,* vol. 14, pp. 20-30.

Chickering, A. W. and Z. F. Gamson (1987). “Seven Principles for Good Practice in Undergraduate Education” *AAHE Bulletin,* vol. 39, no. 7, pp. 3-7.

Cooper, Harris (1989). *Homework* (White Plains, New York: Longman).

Cortright, R. N., Collins, H. L. Rodenbaugh, D. W, and DiCarlo, S. E. (2003), “Student Retention of Course Content is Improved by Collaborative-Group Testing” *Advances in Physiology Education* vol. 27, no.3, September, pp. 102-108.

Crawford, Kathryn, Sue Gordon, Jackie Nicholas, and Michael Prosser (1998), “Qualitatively Different Experiences in Learning Mathematics at University” *Learning and Instruction* vol. 8, no. 5, pp. 455-68.

Edwards, Nelta M. (2007). “Student Self-Grading in Social Statistics” *College Teaching*, vol. 55, no. 2, pp. 72-76.

Geide-Stevenon, Doris (2009). “Does Collecting and Grading Homework Assignments Impact Student Achievement in an Introductory Economics Course?” *Journal of Economics and Economic Education Research* vold. 10. No. 3, pp. 3-14.

Graham, Reginald A. and Beverly L Graham (1997), “Cooperative Learning: The Benefits of Participatory Examinations in Principles of Marketing Classes” *Journal of Education for Business* vol. 72, no. 3, January/February, pp. 149-152.

Grove, W. A., and T. Wasserman (2006). "Homework: Suggestions for Educators" Paper presented at a meeting of the Hutchinson Chapter of Phi Delta Kappa and cited in Geide-Stevenson above.

Hite, Peggy A. (1996), "An Experimental Study of the Effectiveness of Group Exams in an Individual Income Tax Class" *Issues in Accounting Education* vol. 11, no. 1, Spring, pp. 61-76.

Iqbal, Muhammad Javed, Najma Kousar, and Fazalur Rahman (2011) "Collaborative Learning Strategies: Potential Application in Distance Education" *International Journal of Business and Social Science* vol. 2, no. 12 (July), pp. 253-258.

Kirby, N. and C. Downs (2007). "Self-Assessment and the Disadvantaged Student: Potential for Encouraging Self-Regulated Learning" *Assessment & Evaluation in Higher Education*, vol. 32, no. 4, pp. 475-494.

Koppenhaver, G. D. (2006) "Absent and Accounted For: Absenteeism and Cooperative Learning" *Decision Sciences Journal of Innovative Education* vol. 4, no. 1, pp. 29-49.

Leach, Linda (2012). "Optional Self-Assessmet: Some Tensions and Dilemmas" *Assessment & Evaluation in Higher Education*, vol. 37, no. 2 (March), pp. 137-147.

Mallenby D.W and M. L. Mallenby (2003), "Use of Brief Collaborative Quizzes on New Quantitative Lecture Material" *Decision Sciences Journal of Innovative Education* vol. 1, no. 1, February, pp. 141-144

Michaelsen, L., A. Bauman Knight, and L. D. Fink (Eds.) (2004). *Team-Based Learning: A Transformative Use of Small Groups in College Teaching* (Sterling, VA: Stylus Publishing).

Mitri, Michel (2003), "Applying Tacit Knowledge Management Techniques for Performance Assessment" *Computers and Education* vol. 41, no. 2, September, pp. 173-90.

Mumford, Troy V. (2010). "Just Teams: The Relationship Between Team Roles, Fairness and Performance" *Journal of the Academy of Business Education* Vol. 11 (Spring), pp. 12-30.

O'Neil, Harold (2003), "Issues in the Computer-Based Assessment of Collaborative Problem Solving" *Assessment in Education* vol. 10, no. 3, November, pp. 361-74.

Quarstein, Vernon and Polly A. Peterson (2001), "Assessment of Cooperative Learning: A Goal-Criterion Approach" *Innovative Higher Education* vol. 26, no. 1, Fall, pp. 59-77.

Peters, M., B. Kethley, and K. Bullington (2002). "The Relationship Between Homework and Performance in an Introductory Operations Management Course" *Journal of Education for Business* vol. 77, no. 6, pp. 340-344.

Rao, S. P., Collins, H. L., and DiCarlo, S. E., (2002), "Collaborative Testing Enhances Student Learning" *Advances in Physiology Education*, vol. 26, no. 1, March, pp. 37-41.

Sadler, Phillip M. and Eddie Good (2006). “The Impact of Self- and Peer-Grading on Student Learning” *Educational Assessment*, vol. 11, no. 1, pp. 1-31.

Schneider, David I. (2011). *An Introduction of Programming Using Visual Basic 2010.* (Boston: Prentice Hall).

Simkin, Mark G. and William L. Kuechler (2010). “Why is Performance on Multiple-Choice Tests and Constructed Response Tests Not More Closely Related?” *Decision Sciences Journal of Innovative Education* vol. 8, no. 1, pp. 55-73.

Sorcinelli, M. D. (1991). “Research Findings on the Seven Principles” in A. W. Chickering and Z. F. Gamson (Eds.) *Applying the Seven Principles of Good Practice in Undergraduate Education: New Directions for Teaching and Learning 47* (San Francisco, Jossey-Bass).

Spelke, Elizabeth S. (2005). “Sex Differences in Intrinsic Aptitude for Mathematics and Science? “ *American Psychologist* (December), pp. 950-958.

Stearns, Susan A. (1996), “Collaborative Exams as Learning Tools” *College Teaching* vol. 44, no. 3, Summer, pp. 111-112.

Strong, Brent, Mark Davis, and Val Hawks (2004). “Self-Grading in Large General Education Classes” *College Teaching*, vol. 52, no. 2, Spring, pp. 52-57.

Thompson, G., A. Pilgrim, and K. Oliver (2005). “Self-Assessment and Reflective Learning for First-Year Geography Students: A simple Guide or Simply Misguided?” *Journal of Geography in Higher Education* vol. 29, no. 3, pp. 403-420.

Vander Schee, Brian A. (2011). “Let Them Decide: Student Performance and Self Selection of Weights Distribution” *Journal of Education for Business,* vol. 86, no. 6, pp. 13-49.

Vruwink, G. (1998). “Evaluation of Teaching Techniques for Introductory Accounting Courses” *Accounting Review* vol. 62, no. 2, pp. 402-408.

Weems, G. (1998). “The Impact of Homework Collection in Intermediate Algebra” *Research and teaching in Developmental Education* vol. 15, no. 1, pp. 21-26.

Williams, N. (2004), “Daily Shared Quizzes,” *The National Teaching & Learning Forum*, vol. 13, no. 3, (March), pp. 6-8. (Also at http://www.ntlf.com/restricted/v13n3/quizzes.)