Passing the Test: Is Self-Grading a Viable Option at West Point?

Adam J. Kalkstein

This paper was completed and submitted in partial fulfillment of the Master Teacher Program, a 2-year faculty professional development program conducted by the Center for Teaching Excellence, United States Military Academy, West Point, NY, 2011.

Abstract:

Previous research suggests that self-grading, the concept in which students evaluate their own work, has been met with limited success. Researchers have highlighted numerous factors that impact the effectiveness of this assessment method which include: the age and maturity of students, the size of the class, the type of examinations and assignments being graded, and the level of instructor involvement. Most researchers note that students often give self-grading positive marks while reporting increased motivation in the classroom and heightened interest in the subject matter. However, grade inflation remains a common theme, particularly among younger students. Here, a total of 33 West Point cadets are given the opportunity to self-grade a portion of an assignment in an introductory Physical Geography course. Results are mixed, with approximately half the students reporting that they preferred self-grading to a more traditional grading technique.

Background and Literature Review:

A wide-range of research has been conducted on the effectiveness of self-grading, and as a result of varying age groups and methodologies, the results have varied markedly. For example, studies have focused on adult learners (Mueller 1970; Leach 2010), graduate students (Mader 2009), traditional college-age students (Strong et al. 2004; Edwards 2007), and even 7th graders (Sadler and Good 2006). The methodologies employed have been similarly diverse, and this lack of uniformity among the studies has produced highly varied results including students who over-report their grades and others who consistently under-report (Stanton 1978; Davis and Rand 1980). The ensuing literature review will highlight many of the factors that have the largest impact on self-grading, which can then be used to predict the likely effectiveness of self-grading at West Point.

In general, older and more mature students tended to produce lower grades when self-grading, including some adult students who consistently under-reported their grades. For example, Mueller (1970) examined 46 adult students in an introductory Psychology course at the University of Windsor. Incredibly, on a multiple choice exam, 54% of students assigned lower grades to themselves while only 13% incorrectly inflated their grade. Mueller concluded that adult students are harder on themselves and have less confidence in the classroom. One C
student who incorrectly gave herself a D- went so far as to say, “I’m not sure that I even deserve that high a mark.”

Although Stanton (1978) reported similar results among older students, the conclusions were different. In fact, Stanton suggested that older students were not less confident as suggested by Mueller (1970), but rather have a more realistic set of expectations based on increased real-world experience. After additional experimentation that further highlighted the differences between traditional college-age students and adult students, Stanton concluded that it is unfair to use self-grading on traditional-age students since they lack the experience necessary to create accurate grades.

Leach (2010) gave the option to self-grade to 472 mostly adult students, and among these, 120 students (approximately 25%) accepted. Once again, the students assessed themselves rather accurately, with a slight bias towards under-grading (Fig. 1). 32% of respondents accurately graded themselves, 32% over-rated themselves, while the remaining 37% under-rated. Interestingly, better students were more likely to under-rate while worse students tended to over-rate. Similarly, Ellis (2006) hypothesized that more mature students are better able to define their own goals and grade their work against a clear set of standards. On the other hand, less mature students are consistently less comfortable with the self-grading process.

Grade inflation tended to become more of an issue among graduate students, although this phenomenon was more difficult to assess as a result of typically inflated graduate grades (Mader 2009). Possibly as a partial result of the elevated grades, graduate students provided very positive feedback on self-grading. 69% of students reported a higher interest in the material as a result of self-grading. Additionally, 35% reported learning more with only 14% learning less (51% reported learning the same).
Among traditional-aged undergraduate students, self-grading was consistently a popular option, although grade inflation continued to be an issue. Filene (1969) examined nearly 200 students from two history courses at the University of North Carolina. Not surprisingly, while 57% of the students correctly graded their work, 40% of the students reported scores that were too high while only 3% were too low. Acknowledging the grade inflation, Filene suggests that some students mistook enthusiasm for achievement while one student suggested that, “You have a good system but lack a perfect society in which to use it.” Filene hypothesizes that a simple pass-fail structure might lessen the impact of grade inflation among self-grading students.

Numerous studies have highlighted the popularity of self-grading among traditional-age undergraduate students. Davis and Rand (1980) note that while grades were significantly higher for self-graded students (as one example, 62% of students assigned themselves A’s), they were also significantly more satisfied with the course. A research study by Edwards (2007) offered similar results. Among over 100 self-grading students, 86% reported that the system was either “good” or “great” while only 4.3% described it as “bad”, and 10.4% said neither. Students noted that self-grading allows them to identify their own mistakes, and they can learn more as they grade. Additionally, students felt that they gained a better idea of what teachers were looking for, and they very much appreciated the instantaneous feedback. Finally, students also reported that self-grading helped alleviate test anxiety. Edwards does bring up the issue of cheating (a topic that will be addressed later), but hypothesized that it would only be an issue in larger classes.

Strong et al. (2004) examined over 150 students at BYU and noted that 52.4% of students reported that the self-assessment made them more motivated while only 11.7% were less motivated. Considering the importance that motivation plays in education, this is a particularly important finding. Similarly, 66% of students noted that they took more responsibility for their learning compared to only 5.8% who took less responsibility. To summarize, Strong et al. (2004) found that students focused more on what they expect from themselves rather than what teachers expect. Interestingly, students did not feel that their professors were neglecting their duty by not grading assignments; 92.2% of students felt that grading was not a major responsibility of a professor. Unfortunately, grade inflation was a problem. As one example, 93.4% of students acknowledged that average work equated to a “C”, yet when it came time for grading, only 25.7% of those with average work self-graded a C, 60.3% a B, and 11.9% an A. In all, of those who self-graded, 57% received A’s for the course, much higher than the historical average of 31%. Strong et al. (2004) conclude that although grade inflation was a problem, it might very well be a small price to pay. Not only did the majority of students feel that self-grading allowed them to have a better understanding of the material, but a clear majority also thought it made class more enjoyable.

While conducting a self-grading experiment on a younger set of students, Sadler and Good (2006) outline four primary reasons for self-grading. First, they note that self-grading saves instructors a large amount of time and allows for faster feedback to the students. Second, it provides an additional opportunity for students to deepen their understanding of course
material. Third, it allows for a new type of learning that goes beyond simply knowing the answers. Fourth, they conclude that self-learning creates a better classroom experience. In their examination of both peer- and self-grading among 7th grade tests, Sadler and Good note that there is no difference in grading due to gender. However, they find that those who self-grade tend to inflate their grades more frequently while peer-graders have a tendency to under-report grades (Fig. 2). Finally, they conclude that self-grading resulted in increased learning while peer-grading did not.

![Figure 2. Comparison of tallied student grades showing the departure from teacher's grade. From Sadler and Good (2006)](image)

While most scientists agree that self-grading is most effective for older, more mature students, there is still a wide-range of opinion about the effectiveness of self-grading as a whole. Tan (2008) conducted interviews of 16 academics in Australia, and the general conclusion was that self-grading is a vital component for lifelong learning among students. On the other hand, Breidert and Fite (2009) argue the opposite. They suggest that there are too many unknowns that might impact self-assessment, particularly in the military. For example, individual differences, varying methodologies, ambiguity, skill level, and accuracy all impact self-grading. As a result, they conclude that self-grading is not particularly effective, especially in a military setting. The results from the study below echo some of these concerns, as cadets highlight several potential problems with self-grading.
Data and Methods

33 cadets across two introductory Physical Geography courses submitted problem sets covering geomorphology. The first half of each problem set was graded by the instructor and then passed back to the students during the following class. The instructor then informed the students that they would be grading the remaining half of their assignments. The instructor proceeded to review each ungraded question with the cadets, carefully explaining what the correct answers were. Cadets were responsible for grading these questions. After the grading was complete and the problem sets were collected, a brief survey was passed out to the students (Appendix A) to gather their thoughts and opinions on self-grading.

Results and Discussion

Reactions from the 33 cadets were mixed concerning self-grading. 14 cadets (42.4%) preferred self-grading to instructor-grading while 13 cadets (39.4%) did not. 6 cadets (18.2%) reported no difference (Fig. 3). The comments reflected the mixed reaction:

Positive:

“I like being able to write down the correct answers as I grade my own. It helps me figure out why I missed each question.”

“[Self-grading was] more helpful with explanation of what I did wrong.”

“I was forced to closely pay attention to what I got wrong and the corresponding right answer.”

“It helps when an incorrect answer is explained so I understand what I needed in order to receive full credit.”

“A lot of times when I miss points, I’m not positive of how I missed them. By going through each question and seeing your thought process, I know what you’re looking for and can do better on the exam.”

“It was a good way to review.”

Negative:

“I felt kind of guilty because I know what I meant but a grader usually doesn’t.”

“The moral component of grading my own work stressed me out.”

“It’s hard to take points off yourself.”
“Too much of a temptation to inflate grade.”

“I think when people grade their own work, they look for exceptions that will allow them to keep their grade high rather than accept their faults and take the hit.”

Surprisingly, many of the negative viewpoints were a result of morality or a sense of temptation or guilt. This was unexpected, particularly among cadets who are subject to a strict honor code and code of ethics. If something as basic as self-grading a problem set causes such moral angst, these cadets might struggle when faced with more important moral dilemmas while on deployment.

The response was slightly more positive concerning whether or not self-grading impacted cadets’ understanding of material. 16 cadets (48.5%) reported that self-grading allowed them to understand the material more while 17 cadets (51.5%) responded that there was no difference. No cadets believed that self-grading decreased their understanding (Fig. 4). Several cadets commented specifically on their level of understanding based on self-grading:

“I understood it more because I re-read all of the answers I put down as opposed to only re-reading the ones I got wrong.”

“Grading my own work forces me to hear the question and reread my answer: reinforcement.”
“I understood it slightly more, but the way we usually go over the problem sets after you hand them back fully graded is plenty good enough.”

“As long as teacher goes through every question, there is no difference in understanding the material.”

Figure 4. Responses to question: “How did grading your own work impact your understanding of the material (as compared to having me grade it)?”

Conclusions

Previous research has found that self-grading can be an effective method of self-assessment, particularly among older students. Among other things, self-grading can reduce anxiety among students (Edwards 2007), increase student interest in the material (Mader 2009), and save instructors time (Sadler and Good 2006). However, researchers have consistently noted that grade inflation is a problem. Here, 33 West Point cadets were issued a brief survey after self-grading a portion of an assignment. Results were mixed, and there were a similar number of cadets who preferred self-grading versus those who did not. Still, cadets seemed to learn more from self-grading, and this finding is consistent with prior research (e.g. Strong et al. 2004; Ellis 2006). Interestingly, many cadets who did not prefer self-grading highlighted a moral dilemma associated with grading one’s own work. These students cited the conflicting interests of being ethical while trying to achieve a high score.

Tan (2008) argues that self-assessment is a vital component of lifelong learning among students, and results from this study support this conclusion. West Point cadets will be faced with countless moral and ethical challenges in the years ahead. Perhaps an increased level of self-grading at West Point would reinforce right versus wrong and reduce the level of anxiety that cadets might feel when faced with a moral dilemma while serving as an officer in the United States Army.
References


APPENDIX A

BRIEF SURVEY

1) Did you prefer grading your own work?
   Yes  No  No Difference

2) How did grading your own work impact your understanding of the material (as compared to having me grade it)?
   Understood it more  Understood it less  No Difference

3) Please write any comments you have below...Thanks!