USING MASTERY-BASED GRADING TO FACILITATE LEARNING

Robert L. Armacost1 and Julia Pet-Armacost2

Abstract – With a traditional approach to grading, a grade is assigned for a particular evaluation instrument (e.g., quiz, homework assignment, test, project, presentation.) In many cases, the material is reviewed following the grading and the student presumably learns from his or her mistakes. Despite the assumed improved knowledge, the student’s grade still represents what he or she “knew” at the time of the evaluation. This paper describes a concept termed “mastery-based grading” that is intended to actively use the grading system to improve learning. In mastery-based grading, students may re-take examinations as often as they desire in an attempt to improve their grades—ideally, until they have mastered the material. The paper reports the results of the use of the method in two offerings of an undergraduate operations research course. The challenges and benefits of using this approach are discussed and suggestions offered for using the approach in other engineering courses.

Index Terms – Mastery-based grading, course evaluations, learning strategies.

INTRODUCTION

Most engineering courses use a traditional approach to grading where a grade is assigned for a particular evaluation instrument (e.g., quiz, homework assignment, test, project, presentation.) In many cases, the material is reviewed following the grading and the student presumably learns from his or her mistakes. Despite the assumed improved knowledge, the student’s grade represents what he or she “knew” at the time of the evaluation. The overall course grade is typically a composite of these separate evaluations, and therefore, probably not a measure of the level of a student’s knowledge at the completion of the course. Nonetheless, those final course grades are used as an indicator of a student’s achievement in job applications and therefore, probably not a measure of the level of a student’s knowledge at the completion of the course.

For many students, grades are an important motivator. Various models of learning suggest that grades can be used more to increase learning. However, there is little evidence of the effectiveness of different grading approaches. Moreover, it is not clear what the real purpose of grades is—whether they are an indicator of a student’s knowledge, effort, ability, or some combination of all three.

This paper describes a concept termed “mastery-based grading” that is intended to actively use a grading system to improve student learning. In mastery-based grading, students are allowed to re-take examinations as often as they desire in an attempt to improve their grades—ideally, until they have mastered the material. The approach was applied in an undergraduate operations research course required in an industrial engineering curriculum. The paper reports the results of the use of the method in two offerings of the course in separate years. The challenges and benefits of using this approach are discussed in detail and suggestions given for how such an approach can be implemented in other engineering courses.

LEARNING MODELS AND GRADES

We believe that it is generally accepted that a grading system with appropriate feedback to the student is an effective way of increasing the student’s learning. Finding evidence to support that belief is difficult. Our practice as educators with over 48 years of postsecondary teaching experience and extensive interaction with other faculty suggests that this belief is valid. In particular, the practice of reviewing tests in class after they have been graded is intended to help students understand what they did wrong and what the correct approach is. The time spent on this is clearly intended to improve learning. How is learning related to grading?

Sharp, Harb, and Terry [1] described the Kolb learning theory and associated four learning styles (divergers, assimilators, convergers, and accommodators) to develop alternative approaches for teaching writing. Although various approaches to recognize and address different learning styles in engineering education were identified, the consideration of feedback through assessment or grading is not addressed.

A recent study to explore competency gaps in science, mathematics, engineering, and technology (SMET) found the greatest gaps in non-technical competency areas [2]. The study concluded that the critical knowledge, skills, and attitudes to which SMET programs aspire must be student-centered and recommended approaches for student learning, faculty enhancement, and curriculum relevancy. Absent from the discussion was any mention of faculty interaction with students related to grading and feedback.

1 Robert L. Armacost, University Analysis and Planning Support, University of Central Florida, 12424 Research Parkway, Suite 488, Orlando, FL 32826-3207 armacost@mail.ucf.edu
2 Julia Pet-Armacost, Information, Planning, and Assessment, University of Central Florida, 12424 Research Parkway, Suite 225, Orlando, FL 32826-3207 jpetarma@mail.ucf.edu
Moskal, Leydens, and Pavelich [3] focused on understanding the concepts of validity and reliability in engineering assessment, but the role of a grading system and the effect on learning appears to be missing from the discussion. Catalano and Catalano [4] identified seven roles for an instructor to use to move toward a student-centered engineering education. Grading is not considered in this model.

The US Department of Education [5] classifies grading systems into four types:
- Norm-referenced grading systems that assumes students are roughly equal and the goal is to select the best,
- Criterion-referenced grading systems that are based on a fixed numeric scale associated with a letter mark where the scale is invariant to the quality or performance of the students,
- Alternative grading systems that may represent pass-fail structures or non-graded evaluations, and
- Standardized tests that typically generate scores assuming that the scores would be normally distributed.

The general descriptions of the grading systems do not provide any guidance on how the various systems are to be used and their relationship to student learning.

There have been several papers that relate grading systems to providing feedback to students. For example, Pappas and Hendricks [6] developed a “holistic grading” system for use in evaluating writing and communications in engineering. Rather than use a checklist with point assignments for particular elements, they have established a grading rubric that looks at the intended audience and both the content and style of the written or oral communication. Though somewhat more ambiguous, they argue that it is a mechanism for providing stronger feedback to the student.

For an alternate approach, Trevisan, Davis, Calkins, and Gentili developed five design tasks that an instructor can use in order to develop an effective scoring system [7]. They identify some limitations of holistic grading and offer analytic alternatives that provide more feedback to the student.

Shaiewitz [8] made an important distinction between formative assessment and summative assessment in engineering education. Formative assessment in a classroom environment provides information to the instructor that will help to improve student learning. Various classroom assessment techniques such as minute papers, muddiest points, and group problem-solving help to identify areas of emphasis that can result in improved learning. These techniques provide important feedback loops for the instructor. In contrast, summative assessment, directed toward external stakeholders, involves a judgment about the success of the educational process as a whole. Grades typically represent a type of summative assessment.

Hammons and Barnsley [9] provide student-centered guidelines for choosing a grading system. Besides determining whether it should be norm-referenced or criterion-referenced, one should consider

- How the students will view it
- Adequacy of feedback to the students
- Minimum standards of performance, and
- Make-up or retesting possibilities.

In summary, there appears to be little previous work that considers how a grading system can be used to have more of a formative assessment role and improve student learning. In addition, it is not clear that the various approaches to summative assessment in higher education are related to student learning outcomes, and what the final course grades for individual students are intended to represent.

**GRADING AND FORMATIVE ASSESSMENT**

Formative assessment is an evaluation process that seeks to improve the operation that is being assessed. As such, it is the basis for a continuous improvement process. In the context of a classroom, formative assessment seeks to improve student learning and the process continues as learning proceeds.

Grading is generally a type of summative assessment that judges progress or status. A course grade is typically composed of various parts that represent an evaluation of a student’s knowledge with respect to a particular topic. Instructors typically use multiple grading approaches at various times throughout the course. These methods include instruments such as homework assignments, quizzes, tests, projects, presentations, and research reports. The grading method for each instrument is dependent on the type of instrument and the expected content. Instructors typically prefer multiple instruments for a number of reasons.

It is recognized that a true comprehensive evaluation of a student’s knowledge is a sizeable undertaking. Therefore, final exams generally represent a sampling of topics covered, and there is a positive probability that a student will know much of the course material, but be unable to perform on that examination. Therefore, using multiple assessments at various points during a course provides a means of testing more material and minimizing the “bad day” effect. Moreover, using multiple instruments with different approaches allows a student to respond well on those types of instruments that are consistent with his or her learning style. Finally, instructors realize that multiple evaluations throughout the course force the students to keep current with the material and hopefully increase long-term retention of the material.

How can instructors use a grading system as a formative assessment process? What kinds of things can they do? Our experience provides some examples of approaches that instructors often use on a regular basis.

- Drop the low score(s). Frequently, instructors will have a set of quizzes and allow students to drop one or more low scores. This technique is intended to motivate the student who does poorly on a quiz to learn the next set of material well in order to do better on the next quiz.
• Resubmission of corrected quizzes or tests. Another approach is to allow students to resubmit a quiz or test after they have completed the questions (typically outside of class). The student then receives partial credit (typically up to half of the points lost) for the incorrect questions that are now worked correctly. This technique is intended to motivate the student to learn the material so that the question can be answered correctly. The increase in grade provides the motivation. A difficulty is ensuring that the student does his or her own work.

• Resubmission of papers and reports. Unlike quizzes and exams that only require an indication of right or wrong, papers and reports require the instructor to provide constructive feedback to the student. The student may then follow the instructor’s suggestion to improve the content, style, and delivery of the material in the report or paper. There is a significant amount of work in grading the report a second time since it is now a new submission and needs to be evaluated as such. Again, it is typical to allow a student to receive up to half of the points lost on the original submission.

We have used these approaches in courses at the University of Central Florida as well as at Marquette University and the University of Virginia. They have been applied in such courses as statistics, operations research, systems engineering, engineering economy, and introduction to industrial engineering.

Having taken this active approach to using grading as a means to increase learning for many years, we realized, however, that we had not taken the next step that allows the grades awarded on these resubmissions to speak for themselves. If we assume that a student’s grade represents the student’s actual level of knowledge, does it make any sense not to give the student full credit for the resubmission? In practice, we allowed students to recover up to half of the points lost. In reality, we were not using the grades as measure of knowledge, but rather a combined measure of effort throughout the course and knowledge at particular points in time. This realization led to the development of a mastery-based grading system.

MASTERY-BASED GRADING

Mastery-based grading is a system that allows a student to resubmit work as often as desired until the student is satisfied with his or her grade, they cannot improve their understanding of the material, or time runs out in the course. Mastery-based grading can apply to quizzes, exams, projects, papers, presentations, or any other types of evaluation instruments that are used. In this context, evaluation represents a summative assessment at a particular point in time.

In order to improve learning, a resubmission or a retest should force the student to stretch and go beyond the knowledge and understanding expected at an earlier point in time. In practice, this means that each reexamination should be a little bit harder than the previous one. A resubmitted paper or project should have slightly increased expectations with regard to content or style. These expectations need to be clearly articulated to the students prior to their work.

Finally, resubmissions or reexaminations should be timely and follow the previous offering reasonably close in time (perhaps a week later). It is important to ensure that students taking advantage of these learning opportunities are doing their own work. For in class examinations, this is easy to do. For take-home examinations, papers, and reports, it is more difficult to ensure integrity.

We have applied mastery-based grading to part of course evaluation involving in class examinations. In the following sections, we describe the implementation and its results.

MASTERY-BASED GRADING IN OPERATIONS RESEARCH

The Department of Industrial Engineering and Management Systems at the University of Central Florida requires undergraduates to complete a course in Operations Research (ESI 4312) early in the course of study. The course is offered in the fall semester and most of the enrolled students are in their junior year. The objectives of the course follow:

• The primary purpose of this course is to provide the student with a basic understanding of various operations research techniques, their underlying assumptions, the procedures for implementing them, and how to interpret them in the context of operational situations. The course includes coverage of decision-making and model building, optimization of resource allocation problems, project management techniques, and an introduction to descriptive models of waiting lines.

• The mathematical methods include the fundamentals of linear programming including computer based solutions, extensions and special cases of linear programming including transportation and assignment problems, integer programming, networks, queueing theory, and decision theory.

• The emphasis in the course is on the correct use of these tools in industrial engineering and management applications. A reasonable amount of theoretical foundation is required to avoid misapplications and incorrect decision implications. The approach attempts to maintain conceptual rigor while relaxing the mathematical rigor.

Grades are based on student performance on an Operations Research ExplOration report (5 %), four equally weighted mini-cases (6 % each), two term exams (16 % each), a final exam (25 %), and the course case study (14 %). Each of the different instruments is intended to assess particular learning outcomes as the course proceeds.
To implement mastery-based grading, we selected the term exams as the instruments to which the new grading approach would apply. Neither the course case study nor the final exam was eligible to be used because they occurred at the end of the course. The two term exams constituted 32% of the total course grade and satisfy the control issue since any reexams would be given under controlled conditions ensuring individual work.

The other principle involved in a mastery-based grading system is that the content with each successive reexamination is more difficult. In all cases, the same number and type of questions were used. The question content provided the increased difficulty. Figure 1 illustrates two versions of question 13 on two successive reexaminations of the first term exam (about one-third of the way into the semester.).

Exam 1 second reexam
13. (20) You have decided to enter the candy business.
You are considering producing two types of candies: Slugger Candy and Easy Out Candy, both of which consist solely of sugar, nuts, and chocolate. At present, you have in stock 100 oz of sugar, 20 oz of nuts, and 30 oz of chocolate. The mixture used to make Easy Out Candy must contain at least 20% nuts. The mixture used to make Slugger Candy must contain at least 10% nuts and 10% chocolate. Each ounce of Easy Out Candy can be sold for $.25 per ounce and each ounce of Slugger Candy can be sold for $.20 per ounce. Formulate an LP that will help you to maximize your revenue from candy sales.

Exam 1 third reexam
13. (20) Hilltop Coffee manufactures coffee products by blending two types of coffee beans. The Kava beans cost $0.50 per pound and Hilltop has 500 pounds available (already paid for). The Bula beans cost $0.75 per pound and Hilltop has 700 pounds available and does not intend to purchase any additional beans at this time. Consumer quality rating tests, using a 0-100 scale found the Kava beans to have an aroma rating of 75 and a taste rating of 88. The Bula beans have an aroma rating of 82 and a taste rating of 80. Product quality standards for the Hilltop Choice brand require an aroma rating of at least 75 and a taste rating of at least 80. The Hilltop Premium brand requires an aroma rating of at least 78 and a taste rating of at least 84. The marketing department has firm orders for 400 pounds of Hilltop Choice and 300 pounds of Hilltop Premium. Hilltop Choice sells for $3.50 per pound and Hilltop Premium sells for $4.50 per pound. Formulate an LP to develop a production plan to maximize profits for Hilltop.

Another important principle is the timeliness of the reexamination. In all cases, the reexamination was scheduled within one week or less of the previous reexamination. Although students were permitted to take a reexamination at any time, if a student delayed, the reexamination “difficulty” was based on the reexamination that was most recently offered, not on the last one that the individual student completed. The operating effect is that the longer you waited, the more difficult the reexamination became.

Mastery-Based Grading in Operations Research Case Study
Mastery-based grading was used in the undergraduate Operations Research course during the Fall 2000 and the Fall 2001 semesters. The observations include student performance data and responses from an evaluation questionnaire. Students were presented with the opportunity to participate in the mastery-based grading system early in the course. The general demographics of the students are summarized in Table I.

Table I indicates that there was more female participation in both Fall 2000 and Fall 2001. It is interesting to note that there was significantly more participation in the program for both exams in Fall 2000. Not only is participation lower in Fall 2001, there were very few individuals who wanted to take advantage of the opportunity for the first exam. One factor that may have affected this is a change in the university’s grading system between the two years. In Fall 2000, only straight grades (A,B,C,D,F) were used. In Fall 2001, an optional plus and
minus grading system was permitted and that system was used in the Operations Research course.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>MASTERY-BASED GRADING PARTICIPANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Characteristics</td>
<td>Fall 2000</td>
</tr>
<tr>
<td>Enrolled</td>
<td>24</td>
</tr>
<tr>
<td>Males</td>
<td>12</td>
</tr>
<tr>
<td>Females</td>
<td>12</td>
</tr>
<tr>
<td>Participants in Mastery-based Grading</td>
<td>14</td>
</tr>
<tr>
<td>Males</td>
<td>6</td>
</tr>
<tr>
<td>Females</td>
<td>8</td>
</tr>
<tr>
<td>Exam 1—Reexamination 1</td>
<td>11</td>
</tr>
<tr>
<td>Exam 1—Reexamination 2</td>
<td>3</td>
</tr>
<tr>
<td>Exam 1—Reexamination 3</td>
<td>2</td>
</tr>
<tr>
<td>Exam 2—Reexamination 1</td>
<td>9</td>
</tr>
<tr>
<td>Exam 2—Reexamination 2</td>
<td>4</td>
</tr>
</tbody>
</table>

### Student Performance Results

Table II summarizes the results for the individual term examinations and reexaminations throughout the course. It is clear that the test grades improved with each reexamination. The number of students decreased with each subsequent reexamination, but those who persisted found that their grades continued to increase. This result was consistent in both offerings of the course. Some of the increase is likely due to improving test-taking skills, but since there was an increasing difficulty in the test questions, there is a strong possibility that this method is contributing to increased learning.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>MASTERY-BASED GRADING PERFORMANCE RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination</td>
<td>Fall 2000</td>
</tr>
<tr>
<td>Exam 1—All students</td>
<td>81%</td>
</tr>
<tr>
<td>No reexamination</td>
<td>89%</td>
</tr>
<tr>
<td>One or more reexaminations</td>
<td>73%</td>
</tr>
<tr>
<td>Exam 1—Reexamination 1</td>
<td>83%</td>
</tr>
<tr>
<td>Exam 1—Reexamination 2</td>
<td>89%</td>
</tr>
<tr>
<td>Exam 1—Reexamination 3</td>
<td>96%</td>
</tr>
<tr>
<td>Exam 2—All students</td>
<td>85%</td>
</tr>
<tr>
<td>No reexamination</td>
<td>92%</td>
</tr>
<tr>
<td>One or more reexaminations</td>
<td>74%</td>
</tr>
<tr>
<td>Exam 2—Reexamination 1</td>
<td>89%</td>
</tr>
<tr>
<td>Exam 2—Reexamination 2</td>
<td>94%</td>
</tr>
</tbody>
</table>

The students who participated in the Mastery-based Grading system experienced an overall grade increase in both years. A larger percentage of students actually increased their letter grade in Fall 2001 than in Fall 2000. (See Table III.)

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>MASTERY-BASED GRADING EFFECTS ON OVERALL GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination</td>
<td>Fall 2000</td>
</tr>
<tr>
<td>Overall grade increase for participants</td>
<td>3.8%</td>
</tr>
<tr>
<td>Participants with higher letter grade</td>
<td>7</td>
</tr>
<tr>
<td>Participants with no change in letter grade</td>
<td>7</td>
</tr>
</tbody>
</table>

### Student Evaluation of Mastery-Based Grading

Participation in the mastery-based grading system was an individual decision. Because of the experimental nature of the approach, we designed an evaluation of the process that was completed at the end of the semester. A total of 19 students in Fall 2000 and 17 students in Fall 2001 completed the survey. These included 12 from Fall 2000 and 9 from Fall 2001 who participated in the mastery-based grading. The results overwhelmingly supported the use of the system. The following are the responses from all students.

- 68% in Fall 2000 and 76% in Fall 2001 strongly agreed that allowing reexams helps students to learn better. The remainder agreed with the statement.
- 63% in Fall 2000 and 71% in Fall 2001 strongly agreed that allowing reexams is good grading policy. The remainder was neutral or agreed with the statement except for one student in Fall 2001.
- 53% in Fall 2000 and 65% in Fall 2001 strongly disagreed that allowing reexams was unfair to students who did not do so. The remainder was neutral or disagreed with the statement.
- All students indicated that they would like to have the opportunity to take reexaminations in other courses.

Students who participated in the mastery-based grading system were asked another set of questions. The following are their responses.

- 75% in Fall 2000 and 67% in Fall 2001 strongly agreed that they learned better. 25% in Fall 2000 and 22% in Fall 2001 agreed with the statement. One student in Fall 2001 was neutral.
- All students but one in each year disagreed or strongly disagreed with the statement that they only learned how to take the test better.
- All of the students in Fall 2000 and all but three students in Fall 2001 indicated that studying for the reexams did not hurt preparation for other exams and cases in the OR course.
- One student in Fall 2000 indicated that studying for reexams did not hurt preparation for other courses.
- All but two students in Fall 2000 and all of the students in Fall 2001 indicated that the reexams covered the same content and topics.
- Two students in Fall 2000 and one student in Fall 2001 did not agree or strongly agree that the reexams were harder.
- All but one student in Fall 2001 believed that their grade would improve, and all but one of the students in Fall 2001 indicated that they would do it again.

The student responses strongly supported the use of the system. First, there is an overall belief that allowing reexams helps student learning and is a fair system. The students who participated in the system clearly indicated support for it and indicated that they felt that it helped
learning, and was not simply an improvement in test taking. It appears that for the most part, studying for the reexams did not have an adverse effect on other work. Importantly, the students perceived the reexams as covering the same material, but were slightly harder.

LESSONS LEARNED FOR FUTURE USE OF MASTERY-BASED GRADING

There is clear evidence to suggest that this approach to mastery-based grading does improve student learning. In this case, a test is now viewed, not only as an evaluation, but as an assessment instrument that will help a student understand where he or she lacks particular knowledge or capability and provide the student with an opportunity to demonstrate the effects of the additional learning.

The student performance provides the initial evidence for success. Performance on the reexaminations improved in every case. Grades on reexaminations increased by 610 points on average with each additional reexamination. Half of the students in Fall 2000 and 72% of those in Fall 2001 improved their letter grade in the course. That improvement was more likely in Fall 2001 under the plus and minus grading system where the numeric range between successive letter grades was smaller. The student evaluations provide additional evidence of success. Students believed that it was a fair system and that it helped learning.

Using mastery-based grading provides some additional challenges for the instructor. By using the term exams, there was a need to develop a sufficient number of different exams, each with increasing difficulty. During the first offering of the course, those were developed as needed, and then were available during the second offering, requiring only slight modifications to eliminate the benefits to those who might have access to test files. Because of the number of students who were involved, grading the reexaminations was not too burdensome. This could become a major problem in larger course sections.

An additional complication is the scheduling of the reexaminations. Scheduling is complicated by the diverse schedules of the students and the need to obtain a vacant classroom. Doing some preplanning for this eventuality by asking about schedules in advance can be advantageous when trying to make the last minute arrangements.

The implementation of mastery-based grading in an engineering course is relatively straightforward, particularly in the case where analytic examinations play a major role. The use of reexaminations for material that covers 32% of the course grade has more far reaching effects by helping the students to focus on learning rather than testing. They come to understand that a test is an evaluation of their knowledge at a point in time and that they are encouraged to learn more, and can be rewarded for doing so.

The philosophical question that can thwart mastery-based grading involves the question of “what do student grades represent?” Are they measures of knowledge, effort, or what? We believe that grades should be understood as indicators of a level of knowledge at a point in time. Our goal is that all students should have straight A grades. That is the measure of educational success—that all students have the desired level of knowledge. Mastery-based grading can enable such an outcome. If one is concerned about grade inflation or grade distribution, there is probably an underlying belief that students cannot do A level work. In this application, not all students received an A, but over 25% of the students received higher grades than they would have without the mastery-based grading opportunity. Most importantly, they learned more about operations research than they would have without the mastery-based grading opportunity.

CONCLUSIONS

This initial effort suggests that mastery-based grading works to improve student learning. It is a relatively easy process to implement. It does take more time for the instructor to prepare and grade multiple examinations or assignments. But mastery-based grading is an effective way to transform a traditional summative evaluation process into a formative assessment that improves learning.

REFERENCES


