A New Assessment Model in Mechanics of Materials

Ron Averill, Sara Roccabianca, Geoffrey Recktenwald
Department of Mechanical Engineering
Michigan State University
East Lansing, MI 48824
averillr@egr.msu.edu; roccabis@egr.msu.edu; gdr@egr.msu.edu

1. Introduction

The authors have conducted a three-year study to explore the effects of a new assessment model on student outcomes in a sophomore level Mechanics of Materials course. Preliminary results from the first two years were discussed previously [1]. The final set of results and conclusions are presented here.

Based on the results of numerous informal classroom experiments and hundreds of informal discussions with students, it was determined that most students do not use effective study strategies to fully understand key concepts and to master problem solving techniques. Instead, the goal of their current study and test taking strategy is to "maximize partial credit." This strategy works as follows.

1. Memorize problems from the homework, in-class examples, or previous exams.
2. Match each problem on the exam to one of the memorized problems that most closely resembles it.
3. Write down the memorized solution, making adjustments along the way so that the solution looks more relevant to the exam problem.

Because many grading models are based on a poorly defined concept of "correct approach," the above strategy is often very effective at getting a passing grade or better. After all, if the memorized problems have similarities to the exam problems, it is difficult for a grader to determine whether or not a student used the correct approach. When partial credit is then awarded too generously, this flawed study approach is reinforced, and students do not feel the need to solve problems completely or correctly. Note that open-book exams or allowing students to copy solved example problems on personal formula sheets only exacerbates this problem.

Despite being effective at "getting through" a class, this learning strategy has very little value in terms of creating knowledgeable and capable engineers. It is known that pre-test cramming of these example problems into short-term memory promotes almost no retention of the information [2,3]. More importantly, according to studies by cognitive scientists [3-6] (quote below from [3], page 156):

Proceedings of the 2019 ASEE North Central Section Conference

> "Example learners tend to memorize the examples rather than the underlying principles. When they encounter an unfamiliar case, they lack a grasp of the rules needed to classify or solve it, so they generalize from the nearest example they can remember, even if it is not particularly relevant to the new case."

Engineering requires an ability to apply key concepts to a variety of problems that have not been seen before (or memorized). So the approach of maximizing partial credit based on memorizing a few problems is counter to the goals of an engineering education. Furthermore, it can be said that the current partial credit grading model rewards students for pretending that they know how to solve a problem, even when they don't. This means our grading model is promoting behavior that is explicitly unethical for professional engineers, according to the National Society of Professional Engineers (NSPE) Code of Ethics for Engineers [7] (paragraphs II.5.a and III.1.a).

A second practice affecting learning is the copying of homework solutions from online resources. Collaboration on homework has occurred at some level since graded homework was introduced, but the practice of purely copying homework without even thinking about its substance is now so widespread that many have concluded there is no value in assigning course credit for homework. Even online products with randomized parameter values for homework problems have been compromised by shared spreadsheets that allow students to obtain solutions by substituting new values of the parameters. Traditional graded homework is now a high-cost, low-value activity.

In some cases, a passing grade in a class can now be obtained through a combination of copying online solutions to obtain a nearly perfect homework score and maximizing partial credit on exams by memorizing a few example problems. These approaches do not contribute in a positive way to the desired student learning outcomes.

The flawed studying strategies described above are not necessarily new, but their increased magnitude and widespread usage are relatively recent. They are enabled in part by the internet and social media tools, but also by certain types of grading models and course structures. Based on discussions with colleagues at other institutions, these increasing trends are observable across universities and disciplines.

With these ideas in mind, a series of experiments were conducted to measure the effects of a different assessment style on learning. This assessment strategy was implemented in a sophomore level Mechanics of Materials course (i.e., Mechanics of Deformable Solids, ME 222 at Michigan State University). We do not claim that the style of assessment used in the current experiments is unique, but it is uncommon in today's classrooms.

The new assessment model was influenced by a desire to:

- offer very little incentive for using the ineffective memorization strategy described above, while promoting the use of proven learning methods to achieve deeper learning, and
- remove the reward for copying homework solutions, while emphasizing the importance of spaced and varied problem solving repetition (practice) in the mastery of solution processes.

The study results strongly support the hypothesis that assessment style can be used to influence study practices and to promote the desired student learning outcomes. In fact, it may be one of the few ways, if not the only way, to do this.

2. Overview of the Experiments

During the period of fall semester 2016 through fall semester 2018, the authors conducted a series of experiments involving multiple sections of a course in Mechanics of Materials. These experiments are summarized in Table 1.

For each academic year during the period fall semester 2016 through spring semester 2018, two sections of the course used a modified assessment approach (model), while the remaining section (the control) used an assessment approach that mirrors the current standard. Briefly, the standard course design used by Instructor X employed graded homework, two midterm exams, and the use of partial credit based on "correct approach" in the grading of every assignment, while Instructors Y and Z implemented the modified assessment approach. These experiments accounted for the effect of the instructor on the student performance under the new model. The two versions of the new assessment model are described in Section 3.

In fall semester 2018, Instructor X, who previously taught the control sections, adopted the new assessment approach, so that both sections of the course were taught using the same course structure and assessment method.

To measure the effects of the different assessment approaches used during the semester, a common final exam was administered across all sections during each of the semesters. This final exam was developed and graded by the team. It contained different questions each semester, but the exams were consistent with respect to structure, types of problems and level of difficulty. The final exams were not returned to students, so it was presumed that previous versions of the exam were not available to students.

Table 1. Summary of the course models used in each section during the three-semester study. Instances of the new model are shaded, while the control set is not shaded. The number of students in each section is provided in parentheses.

	Fall Semester 2016 (FS16)	Fall Semester 2017 (FS17)	Spring Semester 2018 (SS18)	Fall Semester 2018 (FS18)
Instructor X	Standard model used as control (53 students)	Standard model used as control (107 students)	NA	Version 2 of new model (100 students)
Instructor Y	Version 1 of new model (40 students)	Version 2 of new model (86 students)	NA	Version 2 of new model (91 students)
Instructor Z	Version 1 of new model (60 students)	NA	Version 2 of new model (125 students)	NA

3. New Assessment Strategy and Course Structure

The course design had three primary features: Mastering, Variation, and No Graded Homework.

Mastering

The primary feature of the course design was to move toward a mastery model of grading. In this model, students receive credit only for correct solutions or solutions with minor errors, as described in Table 2. This means that any conceptual mistake made in the solution process results in no credit. The purpose of this grading method is to only give students credit for understanding how to solve a new problem completely, thereby reducing the possible benefits of memorizing example problems.

Table 2. Rubric used to grade each problem on exams.

Competency	Level	Score	Description
Meets Minimum Competency	I	100%	Correct answer fully supported by a complete, rational and easy to follow solution process, including required diagrams and figures
	II	80%	Incorrect answer due to one or two minor errors but supported by a correct solution process as described in Level I
Does Not Meet Minimum Competency	III	0%	Incorrect answer due to conceptual error(s)

In Level II scores described in Table 2, there are two necessary conditions for classifying an error as minor:

1. The mistake is a minor algebraic error, computational error, error in units or significant digits, or other human mistake such as misreading a value in the problem statement.
2. If the identified error had not been made, the final solution would have been correct.

When either of these conditions is not true, the error is assumed to be conceptual and the work does not demonstrate minimum competency. It thus receives no credit.

Grading appeals. In order to reduce the grading effort and increase the benefits associated with the rubric in Table 1, the initial round of exam grading only gave credit to correct solutions (Level I). With complete solutions in hand, students were encouraged to rework exam problems to locate their mistakes. If these mistakes fell into the category described in Level II of Table 1, then a written appeal could be submitted to obtain the partial credit defined in the rubric. More details about the rubric and the grading scheme are described in [8,9].

Locating, classifying and correcting errors on exams can be a very important part of the learning process. This is referred to as *reflection* by cognitive scientists [2], and we prefer that students rather than graders glean this benefit. We hope that this process leads to higher accuracy and grades in the future, all while developing an engineering mindset for checking work and locating mistakes.

Early and Frequent Assessment. In this new course design the timing and frequency of assessment is important. It is recommended that students get two or three early assessments during the first five weeks of the semester. If the assessments are too late in the semester, students will not be able to adjust their study habits in time to complete the course successfully. If the early assessments are too few, students may chalk one or two bad exams up to "a bad day" and not recognize the need to modify their approach. Three exams sends a clear message that what they are currently doing isn't working (or conversely that it is starting to work). Frequent assessment gives students repetition and practice. It is helpful for identifying errors and misconceptions early. It allows the exams to be reasonable in difficulty level.

Multiple exam attempts. We assumed that students had many years of experience working within the paradigm of maximizing partial credit. It could be unreasonable, then, to insert these students into a course that requires mastery. Among other reasons, it is likely that they have not developed proper study habits or the skills necessary to review and correct their work during an examination. To account for this, multiple opportunities were provided on each of the midterm exams. For each of the midterm exams, the final score was the sum of the best scores in each section (described below) from any of the exam attempts. There was only one attempt on the final exam, which had a similar structure as the midterm exams.

In version 1 of the assessment model, three attempts (A, B and C) at each exam were offered. There were four midterm exams, so a total of twelve exams plus the final exam were offered during the semester. With three chances to take each exam, we found that many students did not prepare adequately for the first exam (A), so the effort to create the exam was too high relative to the learning or assessment benefits. Also, twelve exams amounted to an exam almost every week of the semester, which caused a high level of exam fatigue among the students.

In version 2 of the assessment model, only two attempts (A and B) at each exam were offered. The number of midterm exams was increased to five, so a total of ten exams plus the final exam were offered during the semester. The schedule was arranged so that an exam was given in two consecutive weeks, and no exam was given the following week. This repeating pattern of [exam week / exam week / no exam week] provided a mental and emotional break every three weeks, which seemed to alleviate the stress observed in version 1. In addition, two chances to take each exam seemed to be plenty. Against the advice of instructors, some students continued to use the first exam (A) as a practice exam, but this effort was felt to be helpful by both students and instructors.

Reasonable test difficulty and time. Instead of solving a small number of lengthy problems, as was common practice in previous versions of this course, in the current model the exam was divided into 4 sections.

> Section 1: Conceptual questions – Short questions that require little or no calculation but address a fundamental concept in the course.
> Section 2: Simple problems – Questions that require very little time to solve and involve only 1-2 computations. These problems may be very simple cases or address a specific step in the solution to a more complex problem.
> Section 3: Average problems – Medium length problems that contain no avoidable complications.
> Section 4: Challenge problem – One problem that requires a student to demonstrate a complete solution process, often with multiple steps, for a problem that is more complex than those in Section 3.

Additionally, exams were designed with a goal that most students could complete them in about 2/3 of the allotted time. The remaining 1/3 was intended for students to review and correct their work. The total exam time was 90 minutes. For the final exam, the time limit was 120 minutes.

Variation

Each exam was created from scratch by the instructors to ensure that a variety of problems were used. Each version (A, B, C) of an exam covered the same concepts and solution processes, but with a reasonably broad variation in the specific problems that were used in the assessment. The purpose of this variation was to decrease the value of memorizing problems.

As the benefits of problem memorization were removed, students needed to be directed toward a more effective approach to learning. To this end, we created "The Compass." The Compass is a detailed, step-by-step, problem solving process for each type of problem in the course [10]. Throughout the semester, we trained our students to use the Compass to map out their solutions. This is a key part of the current course design, as it gives students a way forward that is productive and helpful.

No Graded Homework

Homework was assigned but not collected or graded. While we did not collect homework, we did strongly emphasize practice. The goal here was to promote meaningful practice, not copying homework solutions for the sake of a homework grade. Students were provided with a list of suggested practice problems, many with fully worked solutions that followed the Compass. These problems were not collected, but students were told that success in the course depended on them using the practice problems correctly to prepare for exams.

We do not suggest that the details of the course design presented here are the ideal or only solutions to the observed student behaviors. Rather, we present them as an attempt to investigate what types of changes in course design might give rise to more beneficial student learning practices as well as improved overall performance.

4. Results

Data were evaluated using one-way ANOVA (analysis of variance) when comparing more than two groups, and using a Student's t-test when comparing two groups. In this section data are presented only as median and mean $\pm$ standard deviation of the mean. Though the ANOVA indicated strong variations among some of the data sets, the details of these variations are not provided here in order to simplify the presentation. The authors felt that these variations, though mathematically significant, were not meaningful in terms of understanding the key effects being investigated.

Comparison of Student Groups Entering the Study

Student records were obtained and analyzed to determine if there were any significant differences among the students in each of the sections as they entered the course. In particular, we measured differences among incoming cumulative GPA's and the grades in the prerequisite course Statics (course code CSE 221 at Michigan State University). A summary of information for each of the three sections is provided in Figures 1 and 2 below. This data shows that, prior to each set of experiments, the variations in the incoming student populations among the various sections were not large enough have a significant impact on the experimental results. The number of credits enrolled in by the students in each Section was also compared and showed no statistical difference (data not shown). The data provided in Figures 1 and 2 are based on a slightly smaller number of students than listed in Table 1 because this data was not available for all students.

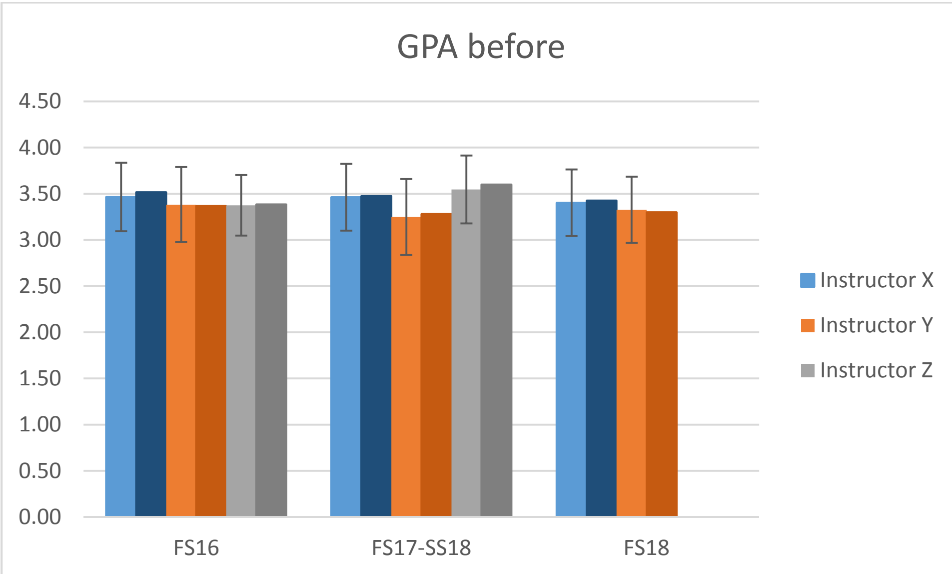

Figure 1. Cumulative GPA (on a 4-point scale) of the students attending Mechanics of Materials prior to entering the course. All data are presented as mean ± standard deviation of the mean (first bar in each set), and median (second bar in each set). Differences are found between Instructor X and Instructor Y (FS17), Instructor Z (SS18) and Instructor Y (FS17), and Instructor Z (SS18) and Instructor Y (FS18) ($p<0.05$).

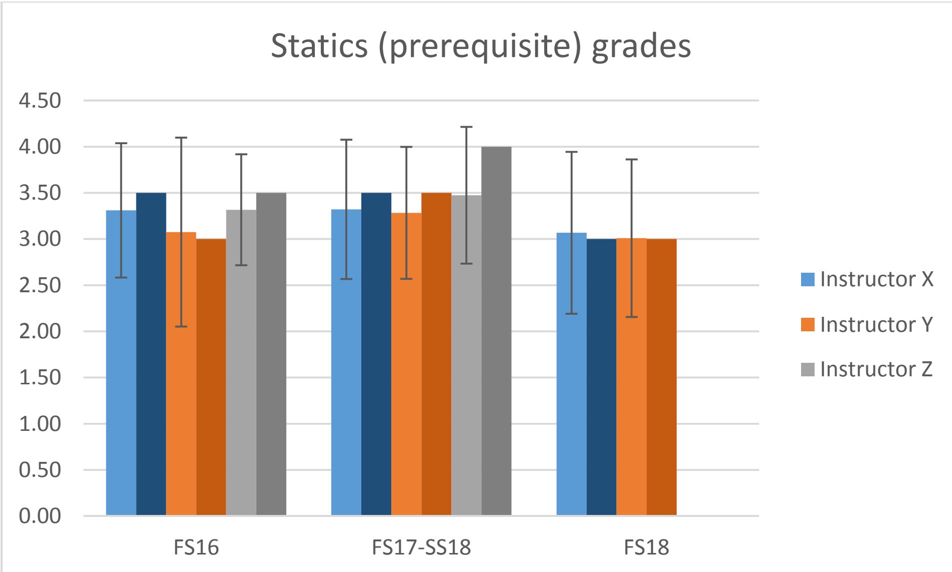

Figure 2. Grades in the prerequisite course Statics for the students attending Mechanics of Materials. All data are presented as mean ± standard deviation of the mean (first bar in each set), and median (second bar in each set). Differences are found between Instructor Z (SS18) and both Instructors X and Y (FS18) ($p<0.05$).

Final Exam Results

The final exam was graded using the rubric in Table 2 for all sections, including the control sections. As described in Table 1, Instructor X taught the control sections during fall semester 2016 (FS16) and fall semester 2017 (FS17). Figure 3 shows the scores achieved in the final exam for each section of the course Mechanics of Materials.

During FS16 and FS17-SS18, the mean and median values of the grades achieved in the final exam by the students attending the control section (Instructor X) were significantly lower than those who attended the sections using the new assessment model (Instructors Y and Z). No statistical difference was found between the scores of Instructors Y and Z. Compared to the control section, the final exam mean scores in the sections using the modified assessment approach were approximately 15 to 30 points higher (out of 100), while the final exam median scores were 21 to 31 points higher. Under commonly used grading levels, this is a difference of 1.5 to 3 letter grades. These results suggest a high impact of the modified course model and grading technique presented here. Further, the data show that equal benefits were obtained for two different instructors.

During FS18, Instructor X adopted the new assessment model, so that both (all) sections of the course used the new model during this semester. The adoption proved to be very successful, with the mean and median scores in the section taught by Instructor X being statistically equivalent to those of Instructor Y during that semester. Considering only the results of Instructor X during the three semesters, the new assessment model resulted in a final exam mean score increase of 19 to 25 points (out of 100), while the median scores increased 21 to 25 points. This is a difference of 2 to 2.5 letter grades.

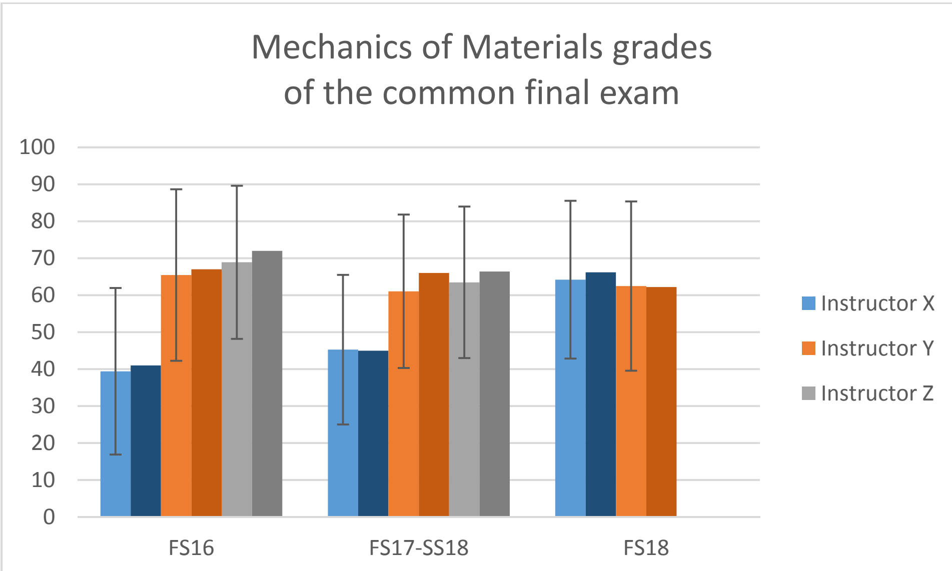

Figure 3. Grades for the final exam of the course Mechanics of Materials. All data are presented as mean ± standard deviation of the mean (first bar in each set), and median (second bar in each set). Differences found between Instructor X (FS16 and FS17) with respect to every other class, namely Instructor Y (FS16, FS17, FS18), Instructor Z (FS16, SS18), and Instructor X (FS18) ($p<0.05$).

5. Additional Observations

When making significant changes to a course design, it is important to take a holistic approach. As the above modified course design was developed, we tried to anticipate issues that might result from any particular change and address them in a way that gave students the best chance of success in the new system.

Our experiences confirm that implementing a stricter partial credit structure by itself would send a shock wave through the system that would certainly cause other problems to flare. This is inadvisable. The various pieces of our implementation worked together to alleviate many of these issues.

The frequency and volume of exams in the proposed approach served as an intense forced practice session. During exams, students would sit uninterrupted for 90 minutes once a week to work on course problems. The repetition and spacing of these sessions likely contributed to improved learning due to increased levels of retrieval practice [2,3,11,12].

The greater amount of class time each week devoted to testing meant that lecture time was significantly reduced under the new model. This made it necessary to focus on the most important topics and example problems in class, and to assign readings and practice problems for out of class work. It is notable that the measured improvements in performance occurred despite the reduction in lecture time compared to the control group.

6. Conclusions

The results of this study strongly support the hypothesis that students will adapt their study habits to changing assessment strategies. Compared to students in a control group that used conventional assessment methods (i.e., partial credit on exams based on "correct approach"), students in sections that used a defined partial credit assessment model scored approximately 1.5 to 3 letter grades higher on a common final exam. This was repeated in multiple semesters with multiple instructors. These results suggest that establishing higher performance standards in conjunction with increasing practice using a compass-guided solution process can lead to significantly higher student performance on exams.

7. References

[1] Ron Averill, Geoffrey Recktenwald, Sara Roccabianca, "Effect of Assessment Methods on Performance in Mechanics of Materials," *2018 ASEE Annual Conference & Exposition*, Salt Lake City, Utah, June 24-27, 2018.
[2] Peter C. Brown, Henry L. Roediger III and Mark A. McDaniel, *Make It Stick: The Science of Successful Learning*, Cambridge, MA: The Belknap Press of Harvard University Press, 2014.
[3] James M. Lang, *Small Teaching: Everyday Lessons from the Science of Teaching*, San Francisco, CA: Jossey-Bass, 2016.
[4] T. Pachur and H. Olsson, "Type of Learning Task Impacts Performance and Strategy Selection in Decision Making," *Cognitive Psychology*, Vol. 65, pp. 207-240, 2012.
[5] M.L. Gick and K.J. Holyoak, "Schema Induction and Analogical Transfer," *Cognitive Psychology*, Vol. 15, pp. 1-38, 1983.
[6] M.A. McDaniel, M.J. Cahill, M. Robbins and C. Wiener, "Individual Differences in Learning and Transfer: Stable Tendencies for Learning Exemplars versus Abstracting Rules," *Journal of Experimental Psychology: General*, Vol. 143, pp. 668-693, 2014.
[7] https://www.nspe.org/resources/ethics/code-ethics
[8] http://newleafedu.com/2018/02/27/grading/
[9] http://newleafedu.com/2018/03/13/rubrics/
[10] http://newleafedu.com/2018/04/17/compass/
[11] H.L. Roediger III and A.C. Butler, "Testing Improves Long-Term Retention in a Simulated Classroom Setting," *European Journal of Cognitive Psychology*, Vol. 19, pp. 514-527, 2007.
[12] B. Rogerson, "Effectiveness of a Daily Class Progress Assessment Technique in Introductory Chemistry," *Journal of Chemical Education*, Vol. 80, pp. 160-164, 2003.