Does a Tailored Course Exam Review Matter?:
Assessing the Effectiveness of Teaching Assessment Tools Used in Conjunction with Exam Review.

Jonathan A. Bodenhamer

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

Senior level Mechanical Engineering courses, by design, cover large subject areas and significant amounts of material. Consequently, many students feel overwhelmed by the exams and are often either unable to create a valid study plan or just chose to inadequately study and risk the consequences. While it is important in these seminal courses to create exams that take students out of their comfort zone and ensure they are truly able to apply the knowledge and engineering tools at hand to solve problems, excessive test anxiety and inadequate instructor guidance can be detrimental to their overall learning experience. ME492, the culminating course in the Automotive Systems concentration of Mechanical Engineering, is just such a course. Three mid-term exams are given during the semester, and there is far more material covered than can be fully included on these tests. The study seeks to provide insight into the following questions:

- Does using Classroom Assessment Techniques to build a short exam review based largely on what the students identify as their primary areas of confusion about the material result in higher grades on the exams?
- Can you create such a review that is effective while still maintaining the rigor of the exams?
- What do students themselves think about the usefulness of such a review?
- What effect, if any, does this have on students’ perception of their own level of knowledge at the end of the course?

This paper provides a qualitative assessment of the effectiveness of using creating and employing an exam review throughout a semester through the use of student test grades, responses to course end survey questions, and an analysis of students’ own feelings about their level of learning at the end of course. The results should be useful to any engineering instructor pondering ways to improve student performance on exams without reducing the challenge of the exercise.
Introduction

Senior level courses in Mechanical Engineering are, by design, exercises in covering large amounts of information about a particular topic area that is normally focused on your concentration area within the major. The intent behind them is to give students added breadth of knowledge in their focus area as well as to give them a feeling for what graduate courses will be like should they choose to continue their educations in the discipline. As course director of ME492: Mechanical Powertrains and Vehicle Dynamics, the significant level of student discomfort with a fast paced, large scope class was very evident. The semester begins with the issue of a textbook, a detailed syllabus, and a large course guide that together provide both a clear road map for the class but also immediately reveal that the entire text and more will be covered in the semester and that there are a large number of graded events, including the dreaded exams. As all educators know, exams covering major portions of the course content, often referred to as “mid terms” as well as a final exam are important to assess the learning progress of the students and also ensure that they have gained sufficient knowledge in different content areas of the class. Despite all efforts to ensure that classroom lectures and board notes covered the key concepts of the test, the results on these mid-term exams from the first semester teaching the course were not commiserate with what was expected, and pondering ways to improve this performance became a focus area for the next iteration of the course. Given this is a senior level class, a dedicated review lesson prior to exams was not conducted, and no real test review was done. The major reasons for this are that there is not time in the syllabus for this and also that as seniors, the students need to be able to prepare themselves for an exam without directed guidance. Consequently, exploring alternative methods to assist students in exam preparation without taking large amounts of class time became a focus. Knowing classroom assessment tools are useful in improving overall teaching, exploration of the hypothesis that using them to create a tailored short review for the students prior to each exam would positively affect their overall performance became the focus of this work.

Why Test

Pondering methods of improving student performance on tests warrants a brief exploration of the overall idea of testing in general. No question, one means of “improving” test performance would be eliminating exams altogether. While many students would certainly appreciate such a thing, there is a large body of research indicating both that exams are a valid way to assess student comprehension and also that some students, based on their learning styles, actually thrive on tests.¹ With a cognizance for the fact that different students have different learning styles, it is possible, and for grading purposes in engineering, critical, that periodic exams be used to assess the learning progress of the students. However, for these to be effective, the instructor must make the exams fair and also ensure that the students are given adequate guidance to avoid a situation where the effectiveness of the exam is lost because nearly all the students “didn’t know beforehand what the prof wanted.”² Should that become the prevailing attitude, very rapidly the students will believe that the tests are impossible and, as a result, exam performance will fall because the perception that studying for exams in the course is futile. However, in engineering, there is a very fine line between giving adequate instructor guidance and telling them directly what is on the exam, the latter option being a major disservice to the students.
because it eliminates the need for engineering students to learn and display the critical ability to solve new problems by using the mathematical tools at hand.

The Syllabus

Considering the potential pitfalls in testing, both from a student and instructor perspective, nearly all courses have one common tool that, at a bare minimum, should give students an idea of the content area of all exams. It is the course syllabus. Distributed on or before the first lesson of a course, this document is a “road map” for the course that should be followed to the greatest extent possible and that students should be encouraged to read and reference on a regular basis. Further, it can and should be a living document, so if changes to the course are needed during the semester, the document should be updated accordingly to reflect reality in the class as the students know it. Adherence to these syllabus policies, as well as including the key concepts for a particular class, will make the document a useful tool for students trying to develop an initial idea of the material they need to study for a particular exam.

Additional Guidance for Exams

In addition to the syllabus, ME492 has historically had a very general study guidance sheet distributed the lesson before each of the three exams given during the semester. Appendix A shows one such sheet. The format is always the same, with the only changes being to the information about the type of question and the point value. These vary with each exam though the total value for each test remains unchanged through the semester. The general premise of the guidance sheets is to re-state pertinent information such as the time and location of the exam, specify when the last available additional instruction (office hours) is available, remind cadets of what resources are allowed on the test, and give very general guidance about what each question will cover. The last part is intentionally vague as the intent is that cadets study and are familiar with all the key material from the particular course block the exam covers so they are prepared for any question that might be asked. Giving more detailed information about the questions would, in the opinion of the instructor, be detrimental for two reasons. The first is that these senior engineering students need to become comfortable preparing for an engineering exam for which they do not know exactly what is going to be asked in advance, and second, because they will certainly not study concepts that are not on the exam if they know they do not have to. The problem with the latter of these two things is that, as specified by the instructor in the course policy letter, all material is testable both on intermediate exams and the Term End Examination, so understanding all the content from a particular block of the semester is critical as they will likely be asked questions on parts of the material during each exam with the Term End used to cover areas not tested previously in the semester. In addition to the guidance sheet shown in Appendix A, old exams are available both on blackboard and in the cadet study rooms. The instructor encourages all students to view old exams to get a feeling for what the test will be like, but also they are warned that only studying old exams will likely result in a poor academic performance as each test is created new every year. The wide scope of the subject area in Mechanical Powertrains and Vehicle Dynamics makes it relatively easy to create tests of very similar difficulty that vary widely year to year. It is likely that part of the overall discomfort with the exams in this course stem from the fact that students viewing old
tests realize that over the years, literally every major concept from the section of the course covered in the exam appears on one test or the other, further reinforcing both the need to study all the material and be very comfortable with the different allowed resources for the test. The final instructor warning on the guidance sheet, *Best of luck….study hard!! If you think you will be able to look everything up and finish on time you might be mistaken!* is intended to drive home the point that historical data in this course shows that students who come totally unprepared for the exams, despite the fact that they are open book and open note, traditionally do very poorly.

**Motivation for Change**

Though the instructor generally considers the exam creation and preparation methodology for this course to be reasonable and overall fair, the reality is that student performance on these events has been trending downward over the past few iterations of the course. Keeping in mind that exams certainly cater to some types of learners, they will remain a fixture in the curriculum of the course. However, further academic research into testing methods and strategies lead the instructor to consider ways to improve student performance without diluting the content of the exams or digressing into a practice of “teaching the test” in class. Davis presents a great deal of material about various ways to reduce student anxiety about and improve their perform on exams. This work details a number of ideas, but two specifically seemed to fit into the construct of this class well and were the basis for the strategy implemented, which were comments on giving students advice on how to study and the direct point of asking students how the instructor can help them feel less anxious about the test. As alluded to by these points, gaining student buy-in to the exam study process, discussing with them what they feel uncomfortable with from the material to be covered on the test, and providing advice on how to study rather than directed areas to focus on would likely improve overall performance while still maintaining the academic rigor of the exam.

**Plan of Action**

Considering advice of experts in the field and tools available to the instructor, developing a means to create a short test review tailored to the needs of the students in general became the focus for improving test performance. After much deliberation, the decision to use various classroom assessment techniques to determine what areas of the material for a particular exam were most confusing to the students, and from that information, create a short exam review, was reached. The overall hypothesis for this review method is that it would satisfy the students’ desire for an exam review and simultaneously act as a sort of discussion with them about the material without being a direct give away for the test content. The review would be created from the areas of most confusion based on the student assessment, independent of what was actually on the exam. Care was taken to avoid intentionally misleading the students into devoting large amounts of time to study material that would not be covered on the exam, but by virtue of the instructor clearly advertising that the review would be based on data collected from the students, they seemed to immediately have a positive feeling about
the plan. It was discussed in conjunction with the distribution of course documents and policy letters on the first lesson of the semester.

Assessment Techniques Employed

Assessing teacher performance through collecting data from students via different classroom assessment techniques is an area of major academic research. In their expansive text on the topic, Angelo and Cross give many ideas and insights into the process of conducting effective classroom research. Considering that the main focus of this assessment was not instructor performance but rather student comprehension of material, or lack there-of, three different techniques were selected from the options listed; the Minute Paper, Muddiest Point Paper, and a tailored variation of the Misconception/Preconception Check. With three exams in the semester, one of these assessment techniques was employed to create the review for each. The instructor considered each of these tools as equally valid for this research, and arbitrarily chose to use the Muddiest Point Paper for Written Partial Review (WPR) 1, the Misconception/Preconception Check for WPR 2, and the Minute Paper for WPR 3. All were administered two lessons prior to the associated WPR so that there was time to create a short, limited to ten minutes, review given at the end of the lesson just prior to the WPR. The Minute Paper and Muddiest Point Paper are well understood and very similar for purposes of this research. The major difference between them is that the Muddiest Point Paper, by its name, directs the students to communicate what concept is most confusing to them and they have open ended time to do this whereas the Minute Paper does not as directly specify what they should communicate and also places them under a time constraint. Observation of student behavior with each assessment technique also was interesting though actions were consistent with what the instructor expected. For WPR 1, the Muddiest Point Paper was used. This was distributed with a few minutes remaining in the class period, and nearly all the cadets immediately took out their course notes and began methodically scanning their material to find the single thing that they felt was most confusing. In many cases the students opted to remain after the end of class to fine tune their answer, and in some cases the instructor was forced to tell them to complete their work in order to allow the follow on section to have seats in the room. For WPR 2, the Misconception/Preconception Check was employed in the form of a five question Blackboard review quiz assigned for completion as homework not later than two lessons before the exam. Again, as expected, nearly all cadets waited to take the quiz until the last possible evening, with four of the twenty-five students not completing the exercise at all. The quiz used is shown in Appendix B. The number of students who failed to complete the quiz at all was consistent with what the instructor expected based on average course behavior to this point in the semester with submission of homework and other graded assignments. For the final mid-term, WPR 3, the Minute Paper was used. As expected, distribution of this created a flurry of frantic searching through notes and attempts by students to rapidly write down information. Many asked for extensions to the minute time, but in order to be consistent with the premise of this assessment technique, the instructor distributed the papers out face down, allowed start work evenly with all participants, and ended the exercise at precisely one minute. The intent was to force the students to list what was really most confusing to them on a quick mental assessment of the material covered since WPR 2 rather than allowing them a lengthy time to ponder their answer.
Exam Review Creation

For each of the above assessment techniques, data from both sections was combined and reviewed by the instructor. As mentioned previously, the only limitation the instructor placed on the review creation was that it would not be deliberately misleading, meaning that even if a large percentage of students listed a topic or subject as a major area of confusion that was not going to appear on the test, it would not be a focus of the review. In the study, this was not a major problem. On the first assessment, the Muddiest Point Paper, a contingent of students did list confusion they still had about a software program called CARSIM that is used in the first lab exercise of the course. Since the test in no way involved use of this program, no review was dedicated to it but the instructor did mention during the review session that a number of students had listed this as an area of confusion and the reason why it was not being covered. In the following two assessments, no such data discrepancy occurred.

Results and Discussion

In order to determine the validity of the hypothesis that tailored reviews based on classroom assessment techniques would produce higher exam scores, it was necessary to compare sets of data with and without benefit of this process. During the first year that the instructor taught ME492, these techniques were not employed. In the following iteration of the course, the techniques were employed. Table 1 shows the student populations from the two semesters of the course.

<table>
<thead>
<tr>
<th>Class</th>
<th># Students</th>
<th>Incoming GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME492 AY 2010</td>
<td>26</td>
<td>3.208</td>
</tr>
<tr>
<td>ME492 AY 2011</td>
<td>25</td>
<td>3.086</td>
</tr>
</tbody>
</table>

Table 1: Comparison of ME492 Courses

It is important, from a data analysis perspective, to note two major aspects of these groups. First, overall size of sample is nearly the same, which is important for accuracy of the data. The second, and even more important point, is that the incoming average GPA for the Academic Year 2010, the group that did not benefit from any tailored WPR reviews, is significantly higher than the incoming average for AY 2011. The difference is 0.14 out of a 4.0 grade point scale, which is significantly higher. Given this, the instructor expected the exam performance of the AY 2010 group to be higher. Figure 1 shows the results of the study.
The results are both interesting and illuminating. Overall, they support the hypothesis, which is encouraging to the instructor. Given the fact that the average academic ability of the AY 2011 cadets is lower than the AY 2010 group, the expected outcome is lower performance on all the exams. However, as show, this is only the case with WPR 1. And, on this exam, the difference in performance was only 1.6% lower for the class. Given that their incoming average GPA is 3.5% lower than their peers in AY 2010, they actually out-performed the AY 2010 group on every exam. Some other interesting, and unexpected, trends appeared as well. The overall difficulty of each exam was almost exactly the same. However, as shown, course performance was strongest on the second WPR and weakest on the last. Nothing in this study offers any illumination on this phenomenon, but the fact that the increase in performance of the AY 2011 group, which benefited from a review, is low along with overall performance may indicate that the overall study time for this exam was lower. Late in the semester, when this WPR is given, workloads are higher throughout all courses due to final projects, papers, and other major projects coming due, creating study time limitations which might be the cause of this. Table 2 lists the average amount of preparation time for each exam reported by cadets.
| Test Time Survey Data  
<table>
<thead>
<tr>
<th>(minutes of preparation time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam</td>
</tr>
<tr>
<td>WPR 1</td>
</tr>
<tr>
<td>WPR 2</td>
</tr>
<tr>
<td>WPR 3</td>
</tr>
</tbody>
</table>

Table 2: Average WPR Preparation Time

However, collected time survey data from the class does not support the theory. In AY 2010, the average study time for WPR 3 was significantly lower than the previous two WPRs. However, for AY 2011, they actually studied, on average, more for each WPR, with their overall average study time being much more consistent exam to exam as well. More research is needed to fully assess the significance of this aspect of the data. The final point of interest based on Figure 1 is that, for WPR 2, the overall performance is the highest and apparently the most beneficial assessment technique was the Misconception/Preconception Check. To draw strong conclusions from this, additional data is needed, but it is possible that the Blackboard quiz used to conduct this assessment, given that there was no time constraint on it, caused students to look harder for the answers and in the process internalize more of the associated information related to the questions. Then, when the in class review was conducted, they were better prepared to benefit from the information by virtue of having more context of the subject material. However, it is possible that this is not a correct notion since there was no point value associated with the Blackboard quiz and no penalty for not completing it, so cadets could put as little or as much effort into it as they wanted. In general, cadets spend little time on exercises that do not directly affect their grades, making it unlikely that they were highly diligent in finding answers to this quiz. Further research into these areas would likely shed further light on these observations and will be conducted at a later date.

Conclusions

Based on the above data, it appears that using various classroom assessment techniques to create a short exam review does improve student performance on the exams. Given that significant care was taken by the instructor to avoid reviewing exactly the material on the exam, in effect “teaching the test,” it is safe to say that improvement is not created by giving the students any unfair advantage. The difficulty of the exams given was very consistent with the exams given to the group that did not benefit from any review, and since the group that had the review was not as high performing academically in general, their superior performance validates the hypothesis. Since the data sample is small for this research, the instructor used course end feedback questions to attempt to further identify if and how the reviews helped. The following question was directly asked of the students in the course end assessment they were tasked to complete.

**E6. Learning Assessment Techniques were used to develop a block review before each WPR this semester. Was this helpful to your preparation for and performance on the WPRs?**
Not all the cadets responded to the question as it is deliberately made voluntary in attempt to collect mostly honest data not given under duress. The responses of interest are listed below.

1. Yes, somewhat
2. Yes, they helped.
3. They were helpful by giving a focus to my WPR preparation.
4. Yes.
5. Yes
6. I believe so. I still feel the WPR guidance is relatively thin, but the notes are great at focusing my thought and studying. However, the conceptual questions and some of the short answer questions are tough, because we cover a lot of the objective data/information through notes but it seems as if some of the more detailed conceptual stuff is left for us to pick out of class discussion... sometimes that is harder to recount and study for than what’s in the notes.
7. no
8. Yes the reviews were very helpful. Sustain
9. Yes
10. Yes
11. Yes the learning assessment helped.
12. yes

As can be seen, of the 12 useful responses, which is approximately half the students in the course, all but one indicated that the reviews were helpful, and some went as far as explaining in detail the ways in which they were beneficial to their test preparation. This data further supports the hypothesis.

In addition to the direct question above, comparative data is taken year after year based on students’ perception of their performance and abilities following the course. Some of this data certainly can be related to the testing process. Figure 2 below shows one such comparative data set.
Course Feedback - ME492 Questions, AY2008 - 2011

E1. I have developed a mastery in driveline fundamental concepts involved in transferring power and torque from the powerplant of a vehicle to the road or terrain.

E2. I can discuss the influence of automotive components on and apply vehicle dynamics to determine vehicle performance in acceleration, road loads, braking, cornering, and ride.

E3. I can apply the basic principles of vehicle dynamics and power trains to a tracked vehicle and evaluate off-road performance of tracked and wheeled vehicles through Terramechanics.

E4. I can use technology to analyze the dynamic performance of an automatic transmission and vehicle dynamics using SIMULINK and CARSIM.

E5. I understand and can discuss current automotive engineering issues and their application to Army technologies and current operating environments and the implications of vehicle technology on society.

Figure 2: ME492 Historical Feedback

What Figure 2 indicates is that student confidence in their overall abilities in the different course learning objectives is higher overall compared to last year, and higher in each category other than question E5. Even in that case, the difference is only a reduction of 0.01. This overall trend of improved student assessment of their own abilities is an indicator that they felt more confident on the WPRs through the semester and have learned more from the course as a whole. Without additional data, it is difficult to state that this trend is a direct result of the exam review plan, but it can be inferred that at the minimum, it was a factor. On broad, difficult exams like the ones in this course, student self-confidence definitely improves overall performance. With all the above data seeming to support the hypothesis, it is important to highlight and discuss one exception. Figure 3 shows this.

Figure 3: Student WPR Course Assessment Question

What is depicted is a comparison between student response to this question for the overall C&ME department (the top red bar), the ME division (the middle green bar), and ME492 (the bottom purple bar). Despite feeling overall confident with their learning, as compared to the division and department, in general students felt the ME492 WPRs were less fair and relevant. From a negative perspective, this data could be inferred to communicate the possibility that the classroom assessment and associated
reviews were not as effective as the instructor had hoped. An alternate view of this data is that it confirms that, despite the review given to the students, the difficulty of the exams was retained and the students remained challenged. Regardless, there was no exam on which the class average was so high as to indicate that the content was too easy or that the review was too direct and gave away the type of problem that would appear on the exam.

Though the data collected is relatively small, all major indicators appear to validate the instructor’s hypothesis that using Classroom Assessment Techniques to create a student driven exam review positively impacts the overall course performance on exams. Further research is needed to draw firm and absolute conclusions, but in this initial study, grades definitely improved while still maintaining the integrity and rigor of the exams, with the added benefit of the students themselves gaining additional confidence in their own level of knowledge following the course. It is a practice to sustain and further study in the future as more data is taken.
Appendix A: ME 492 WPR Guidance Sheet

ME492:

WPR 2 Guidance follows:

**DATE/TIME/LOCATION**

8 October 2010 during class hour (B and C hours)

**ADDITIONAL INSTRUCTION CUTOFF WILL BE Friday, 8 Oct 2010 @ 0830.**

**AUTHORIZED REFERENCES**

- YOUR OWN student notes (no printouts of solutions from past exercises).
- YOUR OWN Textbook (TM 9-8000 and Gillespie). You may WRITE anything in your textbook. You may add external tabs to your textbook.
- YOUR OWN FE Reference Manual
- PERSONAL CALCULATOR

**TEST COMPOSITION**

**PROBLEM #1:** Short answer/True False/Multiple Choice (50 POINTS)

**PROBLEM #2:** Ride (45 POINTS)

**PROBLEM #3:** Handling (65 POINTS)

**PROBLEM #4:** Steering (40 POINTS)

**RECOMMENDATIONS**

1. DO THE READING ASSIGNMENTS.

2. WORK THE HOMEWORK PROBLEMS AND THE WORKSHEETS.

3. USE UNITS APPROPRIATELY – ENSURE UNITS MATCH

4. Lessons 11-18 will be covered.

Best of luck....study hard!! If you think you will be able to look everything up and finish on time you might be mistaken!

MAJ Bodenhamer
Appendix B: WPR 2 Misconception/Preconception Check Blackboard Quiz

Question 1
A vehicle that exhibits Oversteer will have a ______ speed defining the point at which yaw velocity gain and lateral acceleration velocity gain suddenly rise and instability begins.

Question 2
Journey’s Vibration Changes are based on the study of what?
- Human Physiology
- Automobiles and Trucks
- Aircraft
- Ocean-going ships

Question 3
The S-curve describes setting up the front two of a vehicle to be pointed slightly towards each other. It promotes steering stability by making the top sides of the vehicle try to “drive towards” each other, thus keeping the steering wheel centered.
- True
- False

Question 4
Explain what happens in real life at the Damped Natural Frequency you calculate from a vehicle’s 1/4 car model.

Question 5
The damped suspension natural frequency that automotive engineers attempt to design vehicles to have is 1 Hz.
- True
- False
Bibliography


