

Mastery Grading of Engineering Homework Assignments

Jacob Moore
Penn State University
Mont Alto Campus
Mont Alto, PA 17237
jmoore@psu.edu

Abstract— In engineering courses, homework assignments are often intended to be formative assessments where students develop and refine their understanding of the subject material. In reality, however, homework assignments are often evaluated in a way that more closely matches a summative assessment, where the students have little incentive to reflect upon and refine their work. In this paper I discuss a mastery grading system designed to shift the focus of homework assignments from numerical grades to a focus on written feedback and reflection on that feedback. The paper presents the theoretical basis behind the methodology, outlines the mastery grading methodology itself, discusses the results of prior research with the mastery grading system, and presents new results from a two year study examining its effect on student learning, instructor time commitments required for implementation, and student opinions of the method.

Keywords—Homework, Assessment, Mastery Learning

I. MOTIVATION

When developing engineering course materials, most instructors expect the students to continue learning beyond the classroom through the homework assignments they assign. These assignments are not designed to be a final assessment of the students' knowledge; instead, they are designed to be a formative learning activity, where students start to grapple with and apply the concepts and methods they are addressing in lecture. Instructors may offer detailed and helpful feedback to the students while grading these assignments, but it is all too common for students to simply look at the overall grade and then shove the assignments in a folder without even glancing at the written comments.

This points to a disconnect between the motivation of homework assignments where students are still learning the subject and refining their knowledge, and a point-based feedback system that centers on assigning points and judging student performance rather than guiding students to the correct application of their knowledge. In this way, we are treating homework assignments more like summative assessments than the formative assessments they should be.

The goal of the mastery grading system discussed in this paper is to shift the design of the usual homework grading system into something that better matches the goals of formative assessment, and the goal of this study was to evaluate the impact of this new revised design on student

learning, instructor time demands, and student opinions of the homework assignments. To that end, the theory behind good formative assessment design and a review of the relevant prior work is discussed in section II, the final design of the mastery grading system as it was implemented in the classroom is discussed in section III, the data collection methods are discussed in section IV, results are discussed in section V, and the conclusions as well as areas for future work are discussed in section VI.

II. THEORETICAL BASIS AND LITERATURE REVIEW

A. Feedback and Assessment

Assessment in education serves to provide feedback on student understanding to both the instructor and to the student themselves. This feedback can be used primarily to verify that students understand the information at the end of an instructional unit (summative assessment), or it can be used in the midst of an instructional unit to guide students and instructors in future learning (formative assessment) [1]. A comparison of these two types of assessment can be seen in Table I.

TABLE I. FORMATIVE VS. SUMMATIVE ASSESSMENTS

Formative Assessment	Summative Assessment
Measures student understanding in order to adjust instruction and improve student learning	Measures student understanding in order to determine the level of success in reaching the learning objectives of the course
Provided in the midst of the instructional unit	Provided at the conclusion of the instructional unit
Usually constitutes only a small portion of the student's final grade	Usually constitutes a large portion of the student's final grade

Providing timely and high quality feedback to the student through formative assessments has long been known to be a crucial part of education [1]–[3]. If homework assignments are to serve as a type of formative assessment in class, then they must provide both timely and high quality feedback to the students.

In the earlier example where the student simply glances at their grade on the assignment before shoving the assignment

into the back of a folder, there is a breakdown in both quality and timeliness. First, though the instructor may have provided high quality written feedback to the student, the student is focused on the final numerical grade, which does not guide the student on how to correct any deficiencies in their understanding. Second, even if the student does look at the written feedback on homework assignments, there is little incentive to do so until they begin studying for exams, which may be weeks later.

B. Existing Formative Feedback Systems

In order to address the problems described above, a number of approaches have been developed, which can be broadly sorted into three categories: rapid feedback or 'clicker' systems, automated homework assessment systems, and homework or quiz resubmission policies.

Rapid feedback systems generally use multiple choice questions posed by the instructor during class, and require students to use a 'clicker' or other electronic device to answer the questions. The poll results are then presented, providing instantaneous feedback. These systems, particularly when coupled with peer discussion, have been shown to have significant positive impacts on student learning [4]–[6]. Despite the positive results, the systems do have limitations. First, they are designed to be used in class, limiting their usefulness in redesigning the homework assignment, as was the goal of this project. Second, they work best with multiple choice questions and aggregate results, limiting the ability of these systems to offer personalized feedback for complex, multi-step problems.

As an alternative, automated homework assessment systems can also be used to provide instantaneous feedback on student performance outside of the classroom. These systems generally work by having students provide answers to homework questions to a computer based system that then grades their assignment and, for wrong answers, provides instantaneous feedback that may be helpful to the student. Many of these systems allow students to resubmit homework assignments, incentivizing the students to use the feedback to correct misconceptions in their understanding. When properly implemented, these systems also result in significant positive learning gains for students [7], [8]. The downsides to these systems, however, are that they either require a very significant time commitment and programming expertise to develop a quality system, or they require a monetary investment on the part of the student to access commercial tools, a significant factor when considering that college and course textbook costs have been skyrocketing [9].

Finally, resubmission policies with manually generated feedback serve as a low-tech alternative to automated feedback systems. With these methods, instructors provide homework assignments, quizzes, or exams; have the students provide written work for the assignment or exam; grade the assignment or exam, ideally providing detailed written feedback to help the student; and have the student resubmit a corrected assignment or exam to improve their grade. Though these systems do not provide the instantaneous feedback that the automated systems provide, humans have the potential to provide higher quality

feedback. Systems such as this have been evaluated when applied to written feedback for exams [10], [11] and to oral feedback sessions for homework assignments [12]. Though data on student learning is limited in these studies, they do show promise for improving student learning. Drawbacks for these methods include an increased time commitment from the instructor (smaller upfront commitment than automated systems, but larger per-student commitment over the course of the semester), and a delay between student submission and the student receiving feedback. The time required is particularly limiting for large sections, limiting its feasibility in instances where student-to-teacher ratios are large.

III. MASTERY GRADING IMPLEMENTATION

A. Mastery Grading Format

Due to the small class size for the researcher's class sections, the complex and multi-step nature of the problems in the courses, and the desire to keep student costs down, a manual feedback system, from now on simply referred to as the "mastery grading system," was implemented. This system did not adhere to any already established method but instead was designed to maximize the quality and timeliness of the feedback and to incentivize student engagement with that feedback.

The general process employed for the mastery grading system is outlined below.

1. Students are assigned problems from the course textbooks and assignments are due on a roughly weekly basis. Students are expected to follow engineering homework formatting guidelines and show all work.
2. After the problems are collected by the instructor, the instructor 'grades' the assignments giving one of three marks for each of the submitted problems:
 - a. Mastered (M) – indicating that the student has complete and correct understanding of the problem.
 - b. Not Mastered (NM) – indicating that the student has made one or more errors or analysis was incomplete.
 - c. Not Attempted (X) – indicating that the student made no significant effort to complete the problem.

In addition to one of the three marks, students given a 'Not Mastered' mark are given written feedback from the instructor designed to help them correct any errors and guide them to the correct solution.

3. The assignments are returned to the student within a minimum one week timeframe, and from there the students are given one week to resubmit solutions to any problems marked 'Not Mastered'. These resubmissions are required to be stapled to the front of the original assignment, providing a record of feedback and submissions.

4. The assignment resubmissions are regraded using the same method, with new marks replacing old ones, and again returned to the student within a one week time frame. Students are allowed to continue the resubmission cycle as long as needed, though in practice students rarely need more than two resubmissions.
5. At the end of the semester, the student's homework grade is determined by dividing the number of problems the student has mastered (regardless of how many attempts that took) by the number of problems assigned. In this way, students are highly incentivized to continue revising the assignment until they master all topics.

B. Prior Mastery Grading Research

This mastery grading system was first implemented in the fall of 2013 by the author, with another instructor adopting the method in the spring of 2014. Some preliminary data was collected in those spring classes to gather information on student and instructor opinions of the method, student adoption of the method, instructor time demands, and the effect on student grades.

In conducting these studies [13], [14], the authors found that student opinions of the method were very positive, with students indicating they preferred the method to traditional partial credit grading even though it required more effort and time on their part. Additionally, they felt the method was a more fair evaluation of what they knew and the instructors noticed fewer disputes over grades (likely because misunderstandings could always be corrected for full credit).

In examining who took advantage of the resubmissions, it was observed that those at the top, middle, and bottom of the class all took advantage of the method and generally resubmitted until full credit was awarded. This matched the uniformly positive opinions of the method across all levels. This was an encouraging result, since the design was to promote students striving for full mastery of the homework assignments.

In terms of faculty time commitments, there was more time spent grading (in order to deal with the resubmissions), but some time was saved as well since instructors did not need to worry about how many points to take off for mistakes and how to keep that consistent. The net effect was estimated to be about 23% more time spent grading with the mastery grading system when compared to a traditional partial credit system, but results were inconsistent, so more data was to be collected to address this inconsistency.

In terms of the effect on student grades, the all or nothing system resulted in lower initial homework grades, but after allowing for resubmissions, the homework grades for mastery sections were all higher than their equivalent traditional systems with the same instructor and/or course. In fact, because the system allowed mistakes to be corrected, it resulted in many students having perfect grades for the homework portion of their final grade.

Though these studies provided promising results, there were a number of limitations, some of which are addressed in this paper. First and foremost, these studies had no way of experimentally comparing the learning gains, as the method was implemented or not implemented on a section by section basis and no two sections had the same instructor and the same course material. This meant that course differences and instructor differences would be large confounding factors when examining learning gains. The primary goal of this paper is to fill that gap by examining learning gains over a two year period. Second, as mentioned earlier, the instructor time commitments were inconsistent, so further data is being collected to refine that estimate. Finally, though the semi-structured interview allowed researchers to identify some interesting student opinions, some of those results could not be tested for generalizability. This work sought to address this by putting those questions to a broader set of students through a survey.

IV. DATA COLLECTION METHODS

A. Research and Data Collection Goal

As stated earlier, the primary goal of this study was to examine the effect of the mastery grading system on the learning gains of students in order to address a major gap in prior work. Follow-up work was also done to better estimate the instructor time commitments required and the student opinions of the mastery grading system.

B. Study Population

The mastery grading study was implemented in two sophomore level engineering courses (Strengths of Materials and Thermodynamics) at a small public university over two years. These two subjects were purposefully chosen as, prior to the adoption of the mastery grading system, they were found to have very similar average grades. This indicates a similar level of difficulty for both courses.

Class sizes in these sections were small, ranging from three students to thirteen students. During the first year of the study, mastery grading was implemented in Thermodynamics while Strengths of Materials used a traditional partial credit system. For the second year of the study this was switched, with mastery grading used in Strengths of Materials and a traditional partial credit system used in Thermodynamics. A summary of the population can be seen in Table II.

TABLE II. STUDY POPULATION

	Thermodynamics	Strengths of Materials
Year 1	Mastery Grading (n=7)	Traditional Grading (n=13)
Year 2	Traditional Grading (n=3)	Mastery Grading (n=10)

C. Learning Measures

To assess student learning, a number of metrics were used. First, the Strengths of Materials Concept Inventory [15] was used to directly compare students with traditional grading in Strengths of Materials in year one to students with mastery grading in Strengths of Materials in year two. Second, exam performance in the mastery and traditional sections was used as a second point of comparison. As stated earlier, overall grades, including test grades, were observed to have no significant difference in the year before the study. Though different exam questions were used from year to year (in order to discourage cheating), the overall format and style of questions was kept constant. Finally, an additional paired analysis could be done with exam scores for a number of students enrolled in both the mastery and traditional sections at the same time ($n=5$ for year one and $n=3$ for year two).

D. Instructor Time Commitment

To measure the demands on instructor time, the instructor kept detailed records of the time spent grading. During the two year study period, the homework problem sets were not modified but the grading method was from year to year. This means each problem set was graded with each method. To help control for changing class size, a per-student per-problem average time was calculated for each of the four classes under observation. These per-problem averages for the mastery grading sessions included time spent grading originals and all resubmissions for that problem. By comparing the same problem sets with different grading schemes, a more reliable comparison of time requirements could be calculated than was possible in prior research.

E. Student Opinions

To determine student opinions of the mastery grading system, a survey with Likert scale response items was distributed to the students during the last week of class. The survey questions addressed general themes about the mastery grading system as well as specific items previously brought up in prior student interviews. The complete survey can be found in the Appendix.

V. RESULTS AND DISCUSSION

A. Student Learning

The Strengths of Materials Concept Inventory was distributed during the last week of class during each year of the study for extra credit. A summary of results of the concept inventories can be seen in Table III.

TABLE III. STRENGTHS OF MATERIALS CONCEPT INVENTORY SCORES

Year 1 (Traditional)	Year 2 (Mastery)
Average = 55.9%	Average = 64.2%
Standard Deviation = 13.2%	Standard Deviation = 7.4%
Not significant at $P=.05$ (2-Tailed T-Test, unequal variance)	

Though the average score for the mastery grading section is higher, a Student's T-test revealed that the observed differences were not significant at $P=.05$.

When comparing test scores between years, with and without mastery grading in each course, the author also found no significant differences between the two groups in either class. A summary of the exam averages can be seen in Table IV below.

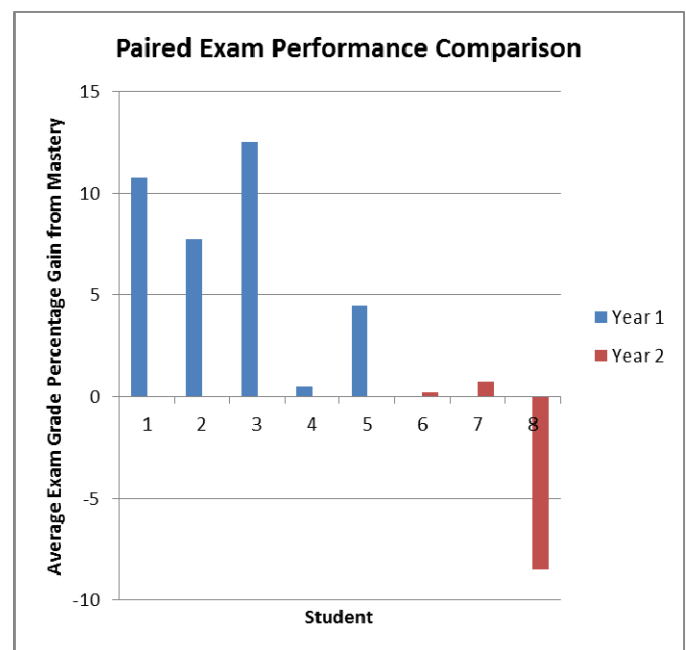
TABLE IV. TEST AVERAGES COMPARISON

	Traditional Grading	Mastery Grading
Strengths of Materials	Average = 83.3% Standard Dev = 6.0%	Average = 85.3% Standard Dev = 8.9%
Thermodynamics	Average = 91.7% Standard Dev = 2.3%	Average = 91.2% Standard Dev = 3.8%

There was also an apparent difference, though not statistically significant at $p=.05$, between exam scores in Strengths of Materials and Thermodynamics overall, contradicting pre-study data.

Finally, a paired comparison of exam scores was conducted for all students enrolled in both the mastery and traditional sections for each year of the study. This compared scores in Strengths of Materials to scores in Thermodynamics, introducing a new confounding variable, but the paired comparison did eliminate student to student differences. A plot of the mastery exam averages minus the traditional exam averages can be seen in Figure 1.

Fig. 1. Paired comparison of exam scores



Again, the observed differences were not statistically significant, and the apparent positive effect of mastery grading in year one and the negative effect in year two can be attributed to the observed differences in Strengths of Materials and Thermodynamics exam averages overall.

B. Instructor Time Commitment

To measure the impact of using mastery grading on instructor time demands, a detailed record of time spent grading was kept by the instructor. Each class used the same problem sets from year to year but alternated the grading method, giving a fairly direct comparison of time requirements. To control for class size, a per-student per-problem average time was used to estimate time demands. This time was calculated by summing up all the time spent grading for the course and then dividing that number by a sum of all the problems the students submitted. This number then represented the average time the instructor spent grading a single problem, including all grading of resubmissions of those problems for the mastery groups. Table V shows the summary of these results.

TABLE V. AVERAGE INSTRUCTOR TIME SPENT GRADING A SINGLE PROBLEM

	Traditional Grading	Mastery Grading
Strengths of Materials	1.00 minutes	2.20 minutes
Thermodynamics	1.45 minutes	1.50 minutes

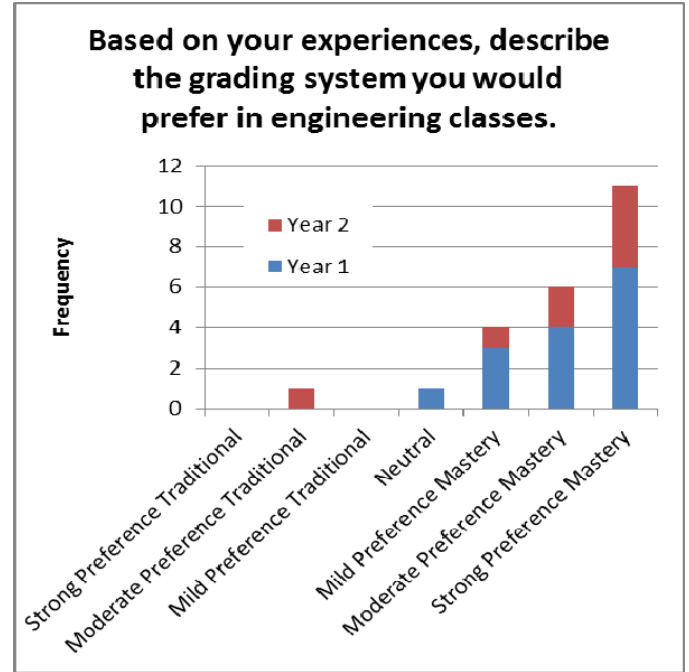
Unfortunately, these results are again fairly inconsistent. The results show a 120% increase in the time required for mastery grading in Strengths of Materials, while only a 3% increase in the time required in mastery grading in Thermodynamics. Averaging these two numbers, we estimate a 62% increase in time spent grading with the mastery grading system, though again the inconsistent results suggest that the findings in other classes and with other instructors may vary significantly.

C. Student Opinions

Student opinions were gathered via surveys distributed to the students in the mastery grading sections each year. The survey consisted of five questions with Likert scale responses as well as one open-ended question. The full survey can be seen in the Appendix.

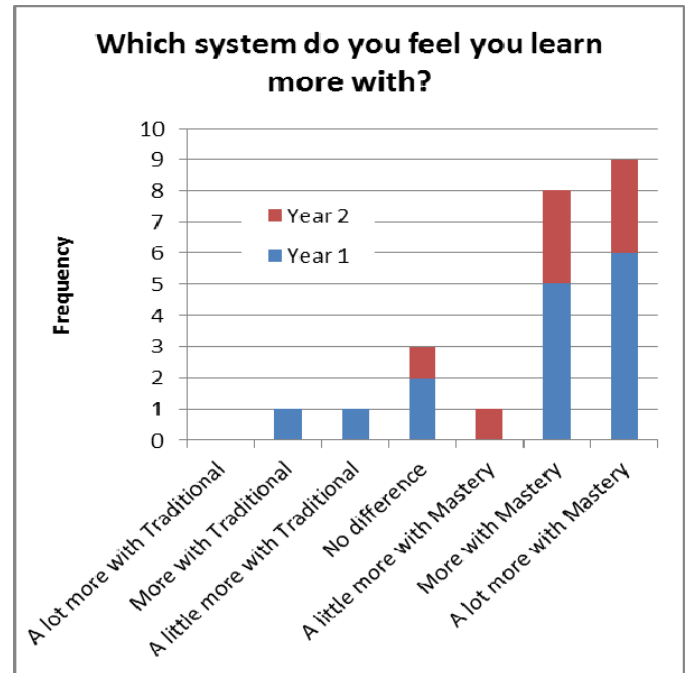
Responses to the first question (shown in Figure 2) addressed students' preference for mastery or traditional grading. As seen in the plot, there is a strong preference for the mastery grading system, with only one student being neutral, one preferring traditional grading and all others preferring the mastery grading system. This matches previous experiences with the system, where student response has been very positive.

Fig. 2. Student survey question one responses



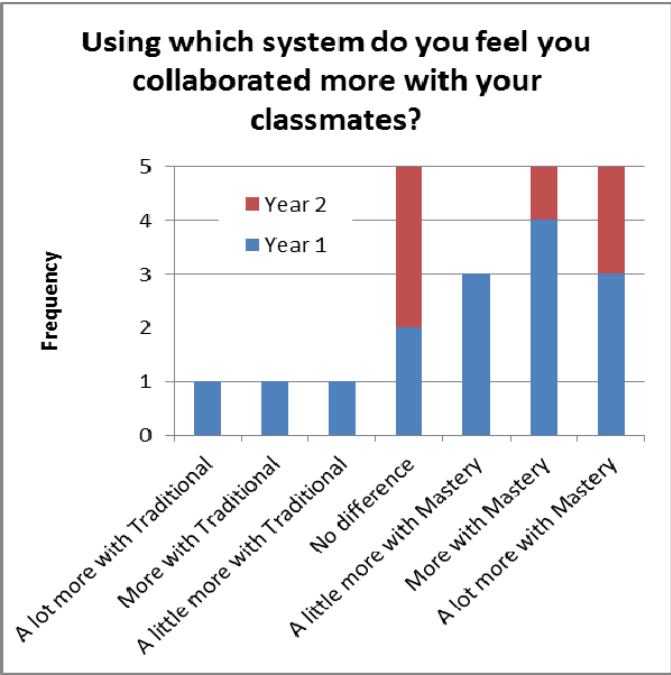
Next, students were asked which system they felt they learned more with. The responses to this question were more varied, as detailed in Figure 3, though most students still felt they learned more with the mastery grading system. Though this is a positive result, this conflicts with the measures of student learning collected and discussed earlier in the paper.

Fig. 3. Student survey question two responses



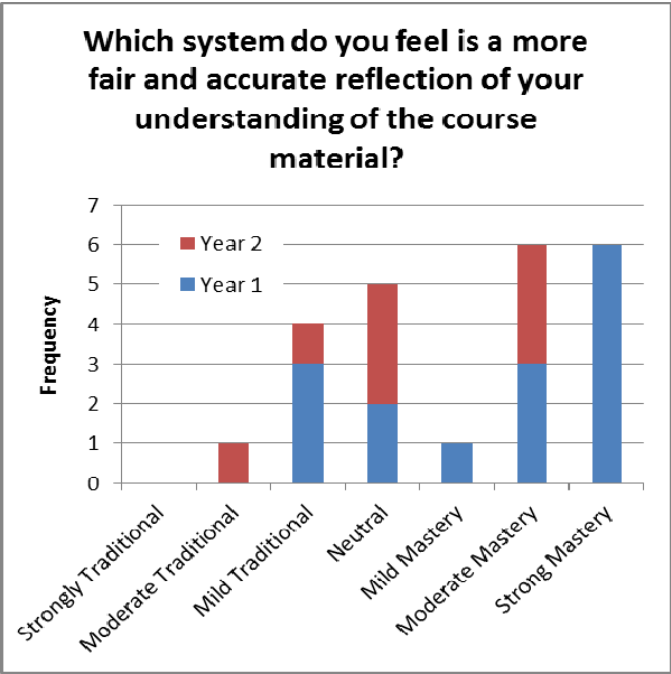
Question three on the student survey asked about student collaboration on homework assignments. Earlier anecdotal evidence from interviews suggested that the mastery grading system encouraged collaboration between students, and although some students did not agree with this, Figure 4 shows that most students did collaborate more with the mastery grading system.

Fig. 4. Student survey question three responses



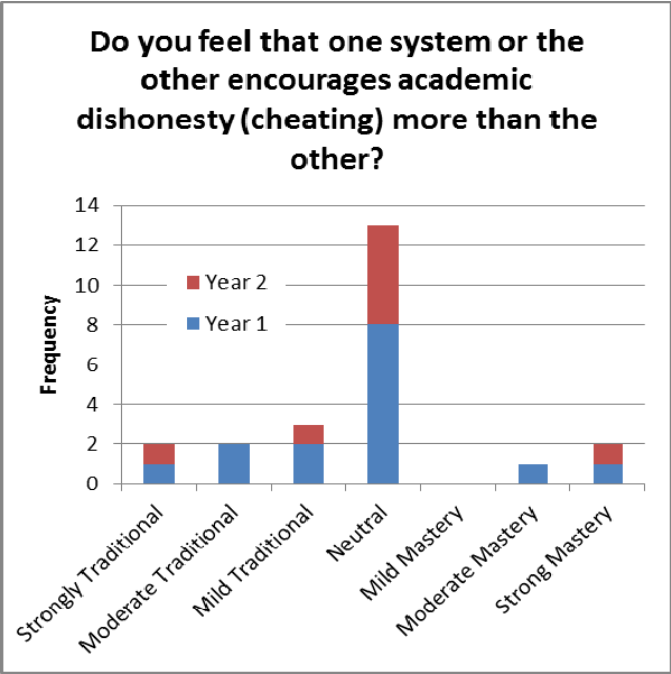
Question four on the student survey addressed how fair and accurate students felt each system was as a grading system. This again went back to comments from earlier interviews, where ‘fairness’ was one of the primary reasons students cited for preferring the mastery grading system. Though the results are not as consistent as the preference in the first question, Figure 5 shows a majority of the students felt that the mastery grading system was a more fair and accurate assessment of their understanding than traditional partial credit grading.

Fig. 5. Student survey question four responses



Finally, due to faculty concerns brought to the researcher’s attention, a question about cheating on homework assignments was added to the survey. Cheating on homework assignments, particularly with student access to solution manuals, is a real problem in engineering courses [16]. Different arguments could be made for why mastery grading encourages or discourages cheating, but from the student perspective as shown in Figure 6, the method does not seem to influence student motivations to cheat or not cheat.

Fig. 6. Student survey question five responses



VI. CONCLUSIONS

Though student opinions of the mastery grading system are very positive, the mastery grading system as detailed here does not seem to have any significant impact on student learning, at least as measured by student performance on the Strengths of Materials Concept Inventory and course exams. Additionally, the mastery grading system does seem to take more time than originally estimated, 62% more time than traditional partial credit grading as estimated in this research, though these results were inconsistent and should be taken with a grain of salt.

In the end, the design of the mastery grading system was successful in its goals of encouraging students to treat homework assignments as more formative assessments and directly address instructor comments, but this did not seem to translate further into observed learning gains. The small sample sizes in this study and the inherent variability and complexity of student learning represent a limitation in the study; however, the manual mastery grading method is also better suited to smaller classes, so a large scale study would be difficult.

The task of providing high quality feedback in a timely manner remains a challenge in education. This study represents one method available to address that need, though more work can clearly be done to improve the feedback systems available.

Appendix – Student Survey Instrument

Thank you for your participation in the mastery grading survey. Your participation will help gather opinions of the implementation of the mastery grading strategy and help improve the method for future years. This survey is anonymous and will in no way affect your course grade, so please be honest in your feedback.

Based on your experiences, describe the grading system you would prefer in engineering classes.

Strong Preference Traditional	Moderate Preference Traditional	Mild Preference Traditional	Neutral	Mild Preference Mastery	Moderate Preference Mastery	Strong Preference Mastery
1	2	3	4	5	6	7

Which system do you feel you learn more with?

A lot more with Traditional	More with Traditional	A little more with Traditional	No difference	A little more with Mastery	More with Mastery	A lot more with Mastery
1	2	3	4	5	6	7

Using which system do you feel you collaborated more with your classmates?

A lot more with Traditional	More with Traditional	A little more with Traditional	No difference	A little more with Mastery	More with Mastery	A lot more with Mastery
1	2	3	4	5	6	7

Which system do you feel is a more fair and accurate reflection of your understanding of the course material?

Strongly Traditional	Moderate Traditional	Mild Traditional	Neutral	Mild Mastery	Moderate Mastery	Strong Mastery
1	2	3	4	5	6	7

Do you feel that one system or the other encourages academic dishonesty (cheating) more than the other?

Strongly Traditional	Moderate Traditional	Mild Traditional	Neutral	Mild Mastery	Moderate Mastery	Strong Mastery
1	2	3	4	5	6	7

Is there anything that you would change about the mastery grading system to make it better?

REFERENCES

- [1] D. R. Sadler, "Formative assessment and the design of instructional systems," *Instr Sci*, vol. 18, no. 2, pp. 119–144, Jun. 1989.
- [2] J. D. Bransford, A. Brown, and R. Cocking, *How People Learn: Brain, Mind, Experience, and School*, Expanded. Washington, D.C.: National Academy Press, 2000.
- [3] E. L. Thorndike, *The Psychology of Learning*. Teachers College, Columbia University, 1913.
- [4] J. C. Chen, D. C. Whittinghill, and J. A. Kadlowec, "Classes That Click: Fast, Rich Feedback to Enhance Student Learning and Satisfaction," *Journal of Engineering Education*, vol. 99, no. 2, pp. 159–168, 2010.
- [5] S. P. Brophy, P. Norris, M. Nichols, and E. D. Jansen, "Development and initial experience with a laptop-based student assessment system to enhance classroom instruction," in *American Society of Engineering Education Annual Conference, Nashville, TN*, 2003.
- [6] S. W. Draper and M. I. Brown, "Increasing interactivity in lectures using an electronic voting system," *Journal of Computer Assisted Learning*, vol. 20, no. 2, pp. 81–94, 2004.
- [7] L. Malmi and A. Korhonen, "Automatic feedback and resubmissions as learning aid," in *IEEE International Conference on Advanced Learning Technologies, 2004. Proceedings*, 2004, pp. 186–190.
- [8] A. Mitrovic, "An intelligent SQL tutor on the web," *International Journal of Artificial Intelligence in Education*, vol. 13, no. 2, pp. 173–197, 2003.
- [9] B. Popken, "College Textbook Prices Have Risen 1,041 Percent Since 1977," *NBC News*, 06-Aug-2015. [Online]. Available: <http://www.nbcnews.com/feature/freshman-year/college-textbook-prices-have-risen-812-percent-1978-n399926>. [Accessed: 19-Apr-2016].
- [10] R. L. Armacost and J. Pet-Armacost, "Using Mastery-Based Grading to Facilitate Learning," in *Frontiers in Education, 2003. FIE 2003 33rd Annual*, 2003, vol. 1, pp. T3A–20–5 Vol.1.
- [11] S. Sangelkar, O. M. Ashour, R. L. Warley, and Oladipo Onipede Jr., "Mastery Learning in Engineering: A Case Study in Statics," presented at the ASEE Annual Conference and Exposition, Indianapolis, IN, 2014.
- [12] J. J. Neubert and J. K. Ness, "The Impact of Cambridge Supervisions on Student Performance in a Dynamics Course," presented at the ASEE Annual Conference and Exposition, Seattle, WA, 2015.
- [13] J. Moore and J. Ranalli, "A Mastery Learning Approach to Engineering Homework Assignments," presented at the ASEE Annual Conference, Seattle, WA, 2015.
- [14] J. Ranalli and J. P. Moore, "New Faculty Experiences with Mastery Grading," presented at the ASEE Annual Conference, Seattle, WA, 2015.
- [15] J. Richardson, P. Steif, J. Morgan, and J. Dantzler, "Development of a concept inventory for strength of materials," in *Frontiers in Education, 2003. FIE 2003 33rd Annual*, 2003, vol. 1, pp. T3D–29–T3D–33 Vol.1.
- [16] J. Widmann, J. Kennedy, and K. Shollenberger, "Student use of author's textbook solution manuals: Effect on student learning of mechanics fundamentals," in *2007 ASEE Annual Conference & Exposition*, 2007.