Determining Optimal Deployment Strategies of MATLAB Autograder to Maximize Student Learning and Engagement

Mr. Jason Scott Smith, Michigan State University
1 Abstract

This work in progress paper will investigate the effect of multiple deployment strategies for MATLAB® Autograder on student performance. Securing the resources required to adequately assess students in large class settings is a common problem faced by many universities and particularly first-year programs, where large lectures are common. These courses require the ability to quickly and efficiently assess and return student work, to optimize learning and satisfy an ever-increasing student desire for instant feedback. The need to achieve these ends in a cost-efficient manner has led to the development and adoption of auto-grading systems in many coding courses. While auto-graders serve as a useful tool in reducing time spent grading by instructors and improving response time for students, it has been noted this advancement can come at the expense of student-instructor interaction and the level of detail in feedback for students. Specifically, auto graders are primarily good at quickly assessing the correctness of answers while not dealing well with nuance or offering students detailed feedback on what specifically they can do differently. Careful development and deployment of assessment strategies is thus critical to maximize student outcomes while minimizing overall cost.

In this work-in-progress study MATLAB® Grader will be used to assess first-year student performance using multiple deployment and grading strategies while student outcomes and attitudes are tracked. Specifically, graded content will be delivered to students using 2 primary modalities. In the first, students will be delivered assignments which will have an “unlimited attempt” grading strategy whereby students are free to repeat their assignments as many times as necessary in order to get the answer.
correct. In the second modality, students will be given an identical set of assignments with a limited number of attempted submissions to the auto-grader. To date, outcomes have been assessed for both student groups through direct comparison of homework grades and through student surveys. In future iterations of this work it is proposed that the results of common examinations also be used to determine which strategy optimizes individual student performance.

2 Background

This study describes the results of student outcomes under varying homework assessment strategies in EGR 102: Introduction to Engineering Modeling. EGR 102 is a freshman laboratory course with 200-350 students per semester, divided into 30 student laboratory groups. Labs meet two days per week, in which students are presented 30 minutes of instruction followed by an assignment. Due to the large number of students and the frequency of assignments, assessment is assisted through the use of Matlab Grader. Matlab Grader is an online product offered by Mathworks that assesses student work by comparing the outputs of student code with solution code. The output of the Grader program is thus an “all or nothing” assessment. This has been used by the instructors as a filter, whereby graders can focus giving feedback solely on sections of code or individuals that are in need of improvement and can automatically bypass correct code. While this method has been helpful in reducing grading time, improved feedback response time, and has been generally well received by student’s, initial deployment of Matlab Grader raised some logistical questions among instructors of EGR 102. Namely, how should the assessments be executed. Initially it was decided that assignments would be assessed in a classical format, i.e. with a single submission. However, due to the frequent occurrence of “technical issues” arising during the submissions process (formatting/saving issues resulting in 0% grades), a two-submission technique was adopted whereby the students could submit their work when completed, see if they made any simple errors, fix them and re-submit for their final grade. However, with this technique an interesting pattern began to arise. The combination of the software’s “all or nothing”
grading technique and the instantaneous feedback given had an interesting effect on student’s who did not initially receive full points. Even though partial credit was given to those who did not complete the problem with 100% accuracy, the fear of instantaneously seeing another “0%” on their screen appeared to discourage students from making further attempts without direct assistance from the instructor. Indeed, the fear immobilized some students to the point that they would not re-submit without verbal assurance from the instructor that the assignment was fully correct. While the assiduousness of such students is commendable, it stands to reason that student autonomy is degraded in such a system and that learning outcomes would thus suffer. In an effort to improve student confidence and engagement, instructors ran a trial wherein half of the class assignments (those given in the first lab of the week) were allowed unlimited submissions. While this achieved the stated goal of improving student autonomy (students felt more confident completing the work themselves through investigation, trial and error), the lack of accountability led to increasingly random student attempts with many students attempting homework through the use of guess and check methods. This lack of a systematic methodology in generating work could also jeopardize long term student outcomes. The purpose of this study is to assess the impact of these and other automatic grader deployment methodologies to determine how best to maximize student learning outcomes.

3 Methods

At this early stage of this work in progress paper (1 semester) a formalized method for creating a controlled study has not yet been deployed. However, initial results from the activities described in the background are available and provide a useful baseline for developing long term methodologies. The data include homework grades, post assignment surveys, and exam grades for students who were subjected to both grading methodologies within the same semester.
3.1 Homework

Homework consisted of short story problems solvable by students using techniques presented in class encoded as either a Matlab script or function. Such homework was given at the conclusion of each laboratory period. Homework at the conclusion of the week’s first laboratory was graded according to the unlimited attempt modality, whereas the second assignment afforded the students only 2 attempts as described in the previous section. This method was the same for all students in the class.

3.2 Surveys

At the conclusion of each assignment, students are asked to complete a reflective activity (survey) that tracks their homework habits, self-assessment of competence, and interest in the material. Three questions from these surveys may help quantify differences in student experience between modalities.

1. How difficult did you find this assignment (Scale of 1-10)?
2. How comfortable would you be using “skill x” in the future?
3. How long did this assignment take you (in minutes)?

3.3 Exams

Students are given 3 examinations throughout the course of the semester. Exams are 80 minutes in length and consist of two to three long form problems solvable by students using techniques presented in class encoded as either a Matlab script or function. At this juncture, exams were not designed with the idea of separating skills learned under the individual grading paradigms. Future semesters will have exam questions dedicated to skills solely learned by one grading method or the other.
4 Results to Date

The following sections outline preliminary results from the first semester of attempting both methodologies.

4.1 Homework

Homework grades for both methodologies were almost identical. Figure 1 shows a bar chart of average grades obtained for homework with each method. Error bars of 1 standard deviation. As would be expected, the unlimited attempts had a slightly higher group average, though not enough to be statistically significant.

![Figure 1: Average homework grades for described methods.](image)

4.2 Surveys

Initial survey results have also yielded slight though not statistically significant differences in student attitudes leaving the assignment which favor the unlimited attempt model. When asked “how confident
do you feel that you could use this skill at a later date” on a scale of 0-10 with 10 being very confident, students responded almost identically between assignments.

![Figure 2: Average student self-assessment when asked their confidence in using the assessed skill at a later date](image)

5 Future Work

Initial results are reflective only of initial instructor observations or student performance in the normal course of teaching the class. At the time of this writing, no controlled studies have taken place. It is the intent of the author to (in the next offering of this course) introduce a side-by-side comparison of techniques, with sections being randomly split into two equal groups. Each group will use one of the grading methodologies exclusively and the results of homework, surveys, and exams compared.