AC 2012-3397: PLAYING THE TA LOTTERY: A STUDY OF HOW TEACHING ASSISTANTS IMPACT GRADES IN ENGINEERING COURSES

Sean Franey, University of Wisconsin, Madison

Sean Franey is a graduate student at the University of Wisconsin, Madison. He is currently working in the PHARM research group under Prof. Mikko Lipasti, joining in Sept. 2010. His research interests include improving the performance and efficiency of data movement in multi-node systems, specifically in the GPGPU realm. His path to UW, Madison, lead through four years on the aircraft carrier USS George Washington in Norfolk, Va., receipt of a B.S. degree from Old Dominion University in computer engineering, and an M.S. degree in electrical engineering from UW, Madison. When not slogging through lines of simulator code, he enjoys brewing award-winning beer.

Anthony Gregerson, University of Wisconsin, Madison

Anthony Gregerson is a Ph.D. student in electrical and computer engineering at the University of Wisconsin, where he recently won the 2012 Exceptional Service Award for teaching assistants. He is a member of the UW's Teaching Academy and the Delta Program in Research, Teaching, and Learning. He has eight years experience teaching as a tutor, Teaching Assistant, and instructor and occasionally writes about testing and assessment for PlusError.com. When not teaching, he designs real-time processing systems for CERN's Large Hadron Collider.

Michael T. Braun, University of Wisconsin, Madison

Michael T. Braun is a Ph.D. candidate in the Department of Communication Arts at the University of Wisconsin, Madison. His research uses advanced statistical methods to look at interpersonal communication technology use through the lifespan.

Playing the TA Lottery

A Study of How Teaching Assistants Impact Grades in Engineering Courses

Abstract – In this paper, we evaluate the impact of Teaching Assistants (TAs) on student grades. We analyze student performance for 6 instances of 2 introductory computer engineering courses, encompassing 12 unique TAs and approximately 800 students. The courses involved both utilize teams of TAs in their instruction, but do so in very different ways. One course relies heavily on TAs in mandatory discussion sections, while the other utilizes TAs to administer lab sections. Within our analysis, we evaluate TA impacts on the grade categories they assign for their own students as well as those grades that are communally or independently assigned and analyze their effect on overall grades. Through our analysis, we determine that those TAs involved in mandatory lab portions of a class have much more impact through their direct assignment of grades, while those TAs involved in mandatory discussion sections have a greater influence on grades that are independently or communally graded. Based on these observations, we discuss how both TAs and instructors can use this information to further the goal of equity among their students.

I. Introduction

Most motivated educators are likely to agree that the individual skills and passion of an instructor can have a major impact on student learning. Likewise, most students are likely to agree that it is easier to get a good grade in a given course with some instructors than with others. With the high-stakes nature of course grades as a factor in scholarships, advancement in higher education, and job opportunities, both students and university administrators have taken a keen interest in monitoring the grade distributions handed out by instructors. Most major universities now publicly disclose grade distributions for courses and/or instructors, and grade information can be an important determiner in which courses or course sections students enroll in[7]. Less attention has been paid to the graduate students working as teaching assistants (TAs) who, though seldom responsible for assigning final student grades, are often involved in critical roles in the instruction and grading processes. In this work, we set out to explore the question of whether differences in individual teaching assistants had a significant effect on students' final grades for a pair of engineering courses taught at the University of Wisconsin – Madison. We were specifically interested in courses that utilize a TA *team* (i.e., multiple TAs contributing to teaching and grading, generally with each TA responsible for one or more *sections* of students) and where students did not have control over which TA was assigned to their class section. Our goal was to determine if this 'TA lottery' gives rise to inequity in students' likelihood of achieving a higher grade in these courses.

Knowledge of this impact is particularly beneficial to the TAs and the instructors themselves. While the coarse-grained view of grade distribution (determined by overall final grade) has a number of stakeholders, the instructors and TAs themselves must take a finer-grained view, consisting of all of the factors that contribute to that final grade and take appropriate action to correct inconsistencies, such as normalizing grades between sections. Before an instructor can make a meaningful decision however, it is crucial to understand the degree to which observed differences in grades between the students of different TAs is based on actual differences in student learning.

Within our study, we break down student performance in a number of categories: grading by a TA, grading by an independent grader, and grading by the TA/instructor team. We evaluate the performance of approximately 800 students over 6 separate instances of 2 introductory computer engineering courses. Within these classes are permutations of 12 unique TAs and 7 unique instructors. From this dataset, we are able to analyze whether factors such as varying experience levels and TA roles (such as discussion mediation versus laboratory instruction) affect different components of their students' grades.

When analyzing our dataset we sought to answer three basic questions: First, do TAs produce statistically significant differences in the grades of their students when compared to other TAs in the same course? Second, if there are differences, what components of the assessment process are affected? Finally, is there a correlation between the amount of previous teaching experience a TA has and the grade outcomes of that TA's students? We hypothesized that this association might manifest itself in either a positive or negative correlation as more experienced TAs might have higher standards for their students but might also be more effective teachers and give more useful feedback on student work.

II. Methods

Course Selection

Our research was conducted within two introductory computer engineering courses at the University of Wisconsin - Madison. We selected courses that used TAs separately as lecture review and discussion leaders ("discussion TAs") and as organizers and supervisors of hands-on laboratory work ("lab TAs"). These two functions are the most common direct interactions between students and TAs in engineering education as noted in the department from which data was gathered for this study. We believe this observation holds at other universities as well.

The first course included in our study is ECE 252, an introduction to many aspects of computer engineering and serves as the first computer engineering course for all students in electrical and computer engineering; and is also part of the computer science curriculum. The course is based on Introduction to Computing Systems by Patt and Patel and covers topics from number systems and combinational and sequential logic design, through assembly-language

programming using a simple ISA[5]. This class has a typical enrollment of 150 to 160 students and is taught in a team-based learning style[4]. In this style, a team of five TAs and the lecture instructor use small groups of students to foster peer-learning with TA guidance. Lectures are held once each week, and material from lecture is then reviewed and reinforced through group exercises in mandatory, twice-per-week, TA-led discussion sections of approximately 15 students each. Within these sections, students often work in small groups of three or four. Graded components include short quizzes and more complex application exercises; TAs are also responsible, as a team, for grading the four exams administered during the semester.

The second course is ECE 352, a continuation of material from the first course, deepening the students' knowledge of combinational and sequential design. This course also has an enrollment of 150 to 160 students and is taught in a more traditional, lecture-focused style with a mandatory lab portion. The class consists of two lecture sections, each with approximately half of the enrolled students, meeting two to three times per week. Material presented in lectures is reinforced with hands-on application through five increasingly complex lab exercises. Completion of these exercises is demonstrated in bi-weekly sections where students program their designs onto FPGA boards and complete a number of in-lab exercises evaluated by their lab TA. The instructional team consists of two lecture instructors, two to three lab TAs, and a discussion TA. Lab TA involvement in this course consists primarily of facilitating two to four demonstration sections and grading pre- and post-lab reports of students in those sections. The pre- and post-lab reports are a significant portion of a student's overall grade (typically weighted as about one third of the overall grade) and consist of schematics, test waveforms, and answers to written questions designed to emphasize particular concepts. TAs, as a team, also grade the three exams administered during the semester.

As we alluded to earlier, there are a number of considerations for both the TAs and instructors with respect to the TA's impact on student grade outcomes. Broadly speaking, we divide these impacts into two categories: direct grade impacts and indirect grade impacts. Direct impacts are TA actions that directly affect their students' grades (e.g., assigning a grade on an assignment that is included in a student's overall grade). Indirect impacts are TA actions that affect their students' performance on other graded material (e.g., providing explanations or examples that improve their students' mastery of course material).

The direct impacts of the TA are related to the nature of the material the TA is responsible for grading and the relative weight this material has in the overall assessment scheme. In a number of courses, grading is distributed among the staff of the course for logistical reasons, particularly in the large classes we are considering, because it is not practical to have every element either independently or collectively graded. When individual TAs assign grades to components that are significant contributors to students' overall course grades, the relative impact that individual TAs have is increased and the effects of inconsistencies in their grading standards are amplified. The indirect impacts on student grades refer to the effect of TA style on student learning. If a TA

provides a different depth or value in feedback or instruction, then that TA's students may be more able to perform on subsequent assignments than students whose TAs have provided less helpful feedback.

Data

We collected student grade data spanning 6 semesters of the courses described previously. This data consisted of written homework grades, exam grades, section homework grades, and overall course grades. In all but one case, an independent grader (an individual that was not also a TA or instructor) evaluated the written homework. In the one case where the written homework grader was not completely independent, this person acted as both the written homework grader and taught 2 sections while the other two TAs taught 4 sections each. Exams were always graded as a team; each grader evaluated one or two specific questions for all students in the class. Section homework consisted of those assignments that were directly graded by a TA for their own students and consisted of lab reports for the lab course and quizzes and applied exercises for the team-based learning course. The overall grade was the weighted average of these three categories.

Prior to analysis, student and TA names were removed and assigned non-identifiable ID numbers (in the case of the students) or letters (in the case of the TAs). No attempt was made to ensure a student had the same number assigned if they appeared in different course instances, but TA letters were kept consistent (e.g., TA A is the same individual for 5 of the different course/semester combinations). This was done to allow us to track TA-specific effects across multiple courses and semesters.

III. Results

Given the exploratory nature of this study, results are considered significant at the $p < .10$ level.

Our first and second research questions sought to assess the degree of variability between TA grading within each course. A one-way ANOVA performed for each class and each semester assessed whether TA was a significant predictor of student score in each of the measured dimensions (e.g., homework grade, exam grade, etc.). For section homework scores, there were significant differences between TAs in 4 of 6 classes. In only one class was there significant variability between TAs in exam grades. In 3 classes, there was significant variability between TAs for final grades. Please see Table 1 for full results.

Our third research question examined whether TA experience had a significant impact on student grades. We analyzed both classes separately, but did not separate by semester. In the introductory course, there were no significant associations between TA experience and student grades. In the more advanced course, there were three significant correlations: TA experience

was negatively correlated with written homework grade (r = -.17, p < .01), section homework grade (r = -.10, p < .05), and calculated final grade (r = -.10, p < .05).

Table 1: One-way ANOVAs assessing differences between TAs across graded components by class and semester

Class/Semester	*Homework Grades*	*Section HW*	*Exam Grade*	*Final Exam*	*Calc Final Grade*	*Adj Final Grade*
252, Fall 2010	F (4, 135) = 3.00**	1.92	2.05*	N/A	2.34*	N/A
Fall 2011	F (4, 146) = 1.99*	4.60***	1.40	N/A	2.73**	2.77**
352, Fall 2008	F (1, 112) = 1.59	13.7****	.482	N/A	3.14*	N/A
Fall 2011	F (2, 123) = .991	.771	1.94	N/A	.536	N/A
Spring 2009	F (2, 142) = .928	8.97****	.451	1.76	.615	N/A
Spring 2011	F (2, 121) = .806	7.37***	.317	N/A	1.02	1.06

Note: * p < .10; ** p < .05; *** p < .01; **** p < .001.

Additional exploratory analyses investigated TA-specific effects across semesters. A one-way ANOVA assessed differences between TAs on final grade. Though the test was significant (F (13, 786) = 2.59, p < .01), conservative Sheffe post-hoc tests failed to reveal any significant differences between TAs. The test was repeated for exam grades. The test was significant (F (13, 786) = 1.54, p < .10, but again no significant differences were found in post-hoc Sheffe tests.

One final exploratory analysis used regression to predict exam grades from section homework grades and TA experience. In the introductory class, the overall model was significant, R^2 = .287, F (2, 288) = 57.9, p < .001. Both homework grades and TA experience were significant predictors. As expected, homework grades were strong predictors of exam grades (B = .516, p < .001); TA experience was also a small but significant predictor (B = .008, p < .05). For the more advanced class, the overall model was significant (R^2 = .249, F (2, 506) = 83.9, p < .001) but only homework scores (B = .424, p < .001) was a significant predictor; TA experience was not significant (B = -.001, p < .921).

IV. Discussion

Our first and second research questions sought to assess the degree of variability between TA grading within each course. To that end, it would appear that there indeed exists variability related to TAs in grading. The fact that significant differences were present in section homework grades but were not present in exam grades is interesting. This suggests to us that the source of these differences is not due to differences in knowledge mastery between students of different sections but may instead arise from some aspect of the grading process. We speculate that some causes are differences in standards between TAs, differences in grading thoroughness, and differences in the level of clarification and assistance TAs provided to members of their section. Further data is necessary to evaluate whether such causes were indeed present.

Furthermore, the fact that the difference in section homework grades contributed so significantly to the final grades, raises concerns that the final grade was not necessarily a complete reflection of a student's mastery of the material, but also a reflection of the relative generosity of their TA. That significant correlations existed in 4 out of 6 instances of the courses we studied, which encompassed 7 of the 12 TAs in the study, further indicates that this is not an isolated incident within the population we studied.

Taken together, along with anecdotal evidence of the largely subjective nature of the lab reports that make up this section homework grade, tighter coordination between TAs may be beneficial. Such coordination may be facilitated by rubrics or other devices to promote greater consistency in grading between sections.

Our third research question wondered if TA experience had a significant impact on student grades. The results for this question would seem to indicate that TA experience can play a factor in student grades, but it is in a direction that one might not expect; we saw negative correlations, indicating greater experience correlates with lower student grades. The fact that this correlation was only present in the more advanced course, where the TAs studied are lab facilitators, confounds the distinction of whether it is the features of the course (complexity, time commitment, etc.) or the role of the TA that drives the correlation. Anecdotally, the more advanced course requires a larger time commitment than the less advanced course.

One line of speculation that can be drawn from this is that as a TA gains experience, any improvement in actual teaching ability may be overcome by some lesser commitment to excellence in teaching. Perhaps completion of multiple semesters of teaching wears them down. Where they once entered the semester full of excitement (and/or apprehension) of the unknown and a desire to be outstanding (or a fear of being mediocre), they eventually enter the semester with less focus on their teaching duties or demotivation due to greater knowledge of the work that will be required. It may be easier to maintain a high level of excitement and devotion to something you do not expect to consume a significant chunk of your time, while it is harder to maintain the same level of motivation for something you feel certain will take away too much time from other duties. Another, non-mutually-exclusive explanation could be that TAs learn what they can "get away with." In other words, a TA decides which elements of teaching they wish to focus on and devotes less attention to other areas that they deem to be less important, less interesting, etc. Of course all of this is pure speculation informed by anecdotal evidence, since our data has no measure of TA motivation, intentions, or desires.

Our exploratory analyses initially investigated whether differences between TAs could be explained separate from semester and class effects. The results from this test are interesting in that they do reveal significant difference between TAs when compared to grades they either have limited direct impact on (final assigned grades) or none at all (exams). The fact that more conservative tests reveal no significant difference however, indicate the need for more data to

confidently attribute these differences to TAs and not other factors such as time of the offering, which could affect student distribution due to some conflicting with other courses or otherwise being offered at times that are more convenient for a specific demographic of students (e.g., evening courses).

Our final exploratory analysis using regression to predict exam grades from section homework grades and TA experience revealed two very interesting results. On one hand, homework grade is a very strong indicator of exam grade. This is not surprising; good students perform well on both homework and exams. What is more interesting is that outside of this affect, TA experience remains a significant predictor. While it is very minor, the fact that a more rigorous analysis of the data – controlling for homework grades – still reveals an association between TA, experience, and grade lends credence to the belief that a TA *does* impact a student's grade. This analysis serves to further strengthen the claim that students may be receiving different outcomes based on their TA. Fortunately, this analysis reveals an association with exam grade and is therefore more likely to reflect a genuine increase in student knowledge and not merely an artificial inflation. While this conclusion is reassuring to imply that the sanctity of our grade assignments is maintained, a new issue is presented that some students may be receiving higher quality instructions than others. While this is an issue that may not have a complete solution, it is important to acknowledge the potential for difference and strive to minimize it.

Directions for Future Research

Given the exploratory nature of this study and the results obtained so far, there are a number of avenues for future research. First, it would be interesting to know what traits of a TA – beyond experience –affect their quality of instruction (indirect effects) and their grading accuracy and consistency (direct effects). While our data has the ability to evaluate the trait of experience, it lacks the traits of motivation, commitment to teaching, time expended on grading, or – more abstractly – ability, among others. While many of these traits are difficult to objectively measure, one can obtain a proxy for certain characteristics from existing data. Teaching evaluations, for instance may be interesting to correlate with student achievement. Also, monitoring how teaching evaluations changed with experience, and how TAs with similar evaluations but different levels of experience compare with respect to student achievement (i.e. how TAs that just have "the right stuff" off the bat compare to TAs that had to gain experience to become good) could reveal interesting information.

Second, it would be interesting to know what factors of the advanced course led to the negative correlation of experience and grades. While "motivation" may be a difficult trait to isolate, one could design an experiment to evaluate the correlation of TA experience and grades on a more advanced course, or one with similar time commitment, or one with similar roles. Of course, the more of these dimensions that could be gathered, the more firmly conclusions could be drawn.

Third, it would be worthwhile to test whether some of the techniques we proposed for mitigating variation in grading of section-specific assignments reduced the effect of TA section on student grades.

V. Conclusion

In summary, our data and analyses support the claim that TAs have an impact on student grades in the courses used in our study. While unable to make definitive conclusions about the causes of these differences, these results can be used to direct attention to certain aspects in relation to teaching assistants and grades to help mitigate potential issues.

It is important to note that our results were based on a population that was limited to two courses at one university and therefore one should be careful about trying to generalize these results to all university classes. That said, we believe that the potential issues highlighted in this study are of interest to a broad audience of instructors, and that all instructors should be aware of the potential TA-related effects on student grades in their courses. The existence of either direct or indirect grading effects presents problems of inequity for students. Course grades are high-stakes assessments, influencing decisions important to students' futures such as scholarships, job applications, and admission to advanced study programs. Instructors therefore have a duty to their students to ensure that the grades they receive are as fair and accurate as possible and that no students are given unfair advantages due to the TA lottery.

Although error exists in all forms of assessment, there are steps instructors can take to mitigate TA influences on assessment accuracy. In our studies, significant differences primarily occurred in homework that was graded separately by each TA and only rarely occurred in exams that were graded in a communal, uniform manner. Although it may not be logistically possible to use communal grading for all items, it might be possible to adopt practices to make individual grading more uniform, such as the adoption of communally designed rubrics, instructor-led grading tutorial sessions, or TA peer review. When possible, instructors might also try to reassign TAs to different sections over the course of the semester, although this would likely pose scheduling challenges.

We hope the results of this exploratory study have highlighted the need for further research into this issue. Although we were able to uncover interesting correlations in the courses we studied, the limitations of our data set did not allow us to make confident statements about the causes of correlations between TAs and student grade outcomes. We encourage instructors of multi-TA courses to conduct similar analyses to determine whether differences in TAs produce significant differences in student outcomes for their courses and identify means of minimizing such effects.

References

1. Carroll, J. G. (1980). Effects of training programs for university teaching assistants: A review of empirical research. The Journal of Higher Education, , 167-183.

2. Hoffmann, F., & Oreopoulos, P. (2006). Professor Qualities and Student Achievement,

3. Jacobs, L. C., & Friedman, C. B. (1988). Student achievement under foreign teaching associates compared with native teaching associates. The Journal of Higher Education, , 551-563.

4. Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2002). Team-based learning: A transformative use of small groups Praeger Publishers.

5. Patt, Y. N., & Patel, S. J. (2003). Introduction to computing systems: From bits and gates to C and beyond McGraw-Hill Higher Education.

6. Prieto, L. R., & Altmaier, E. M. (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. Research in Higher Education, 35(4), 481-497.

7. Sabot, R., & Wakeman-Linn, J. (1991). Grade inflation and course choice. The Journal of Economic Perspectives, 5(1), 159-170.