2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51743

Work in Progress: Training GTAs in Feedback and Grading Practices for a Mechanical Engineering Lab Course

Erin Elizabeth Egersheim, University of Connecticut

Erin Egersheim has received a bachelor's degree in biomedical engineering from The Ohio State University and a master's degree in biomedical engineering from the University of Connecticut. She is currently pursuing my Ph.D. in engineering education at the University of Connecticut.

Martin Huber, University of Connecticut

Work in Progress: Training GTAs in Feedback and Grading Practices for a Mechanical Engineering Lab Course

1. Introduction

In graduate programs, students have to juggle many responsibilities in their pursuit of their advanced degrees. While coursework and, for some, research, are graduate students' primary academic priorities, an increasing number also work as graduate teaching assistants (GTAs) [1, 2]. For many of these students, their GTA position is their first encounter instructing undergraduates. During this period, GTAs are supervised by the course's instructor of record (IOR), who are often faculty members. Depending on the course and the number of enrolled students, a GTA's responsibilities can range from grading to instruction in lecture or laboratory settings [3, 4]. Fundamentally and in name, GTAs function as assistants to the IOR to help ensure that the course operates effectively and that student learning objectives are met.

However, unlike faculty, access to professional development for GTAs is often limited to generalized programming delivered at the start of their appointment [5]. Despite the limited access, participation in directed professional development has shown to improve not only the performance of the GTA but also the outcomes of the students they support [6, 7]. With substantial variability across courses, professional development tailored to individual GTA appointments can be difficult to develop and implement at scale. As a result, IORs are frequently responsible for designing and implementing their own training for teaching assistants, which, depending on its structure and intent, can vary widely in effectiveness [8].

The effectiveness of professional development is not only critical for providing expectations and ensuring preparedness but also plays a role in a GTA's belief in their teaching abilities. Higher self-efficacy, or in context to instructors, teacher self-efficacy, is often associated with improved performance and increased confidence in the individual's role [9–12]. Research has demonstrated that students are perceptive of their instructors' self-efficacy and tend to respond more positively to instructors who exhibit higher levels of teacher self-efficacy compared to those with lower levels [7, 13, 14]. While teaching experience can contribute to increased self-efficacy, prior work suggests that the addition of professional development more effectively improves self-efficacy [11, 15, 16],which is particularly important for GTAs who often begin their appointments with limited teaching experience.

The pilot of this study will be conducted in a mechanical engineering laboratory course in which lab reports constitute the majority of graded assignments. The structure formal training designed for this study reflects the principles of the Cognitive Apprenticeship Model (CAM), which emphasizes learning through modeling, guided practice, and reflection [17]. Through the formal training the GTAs will learn about the assessment style which has been based on traditional feedback used in writing instruction. Courses in STEM, including engineering, have successfully incorporated similar approaches through Writing Across the Curriculum initiatives that highlight writing skills, the writing process, and program-specific writing style [18, 19]. We aim to determine whether this training structure is effective in improving teaching self-efficacy, grading consistency, and ability among GTAs. If shown to be effective, this training model may support expansion across departments. However, it is acknowledged that this pilot involves a limited sample and cannot confirm broadly generalizable outcomes.

2. Methods

2.1 Participants

The participants consisted of the eight GTAs assigned to the third-year mechanical engineering lab course at a large Northeast public university during the study period. However, one participant only completed the post-training survey and grading practice; therefore, data from this participant were not included as no comparison could be made. The GTAs were assigned to this course according to departmental procedure, and this assignment occurred without the input of researchers. Of the GTAs who provided demographic information, only two of the seven GTAs indicated that they had previous experience grading writing. One undergraduate teaching assistant (UTA) was hired after the training session to grade reports but was excluded from the study; they were hired after the training session and therefore did not receive the intervention.

2.2 Training Design

All GTAs for this course were required to participate in multiple meetings prior to the first laboratory session to learn about course objectives, lab management, and lab procedures. The structure of the training aligns with key elements of CAM [17] which emphasize modeling and coaching through the use of examples of grading, as well as guided practice and reflection through opportunities to grade example reports and group discussion. To develop the training content, the lead researcher collaborated with IOR, the Center for the Integration of Research, Teaching, and Learning, and the university's writing center. Additionally, the subjective nature of writing was considered, as variability in how graders interpret writing quality, apply rubric criteria, and assign partial credit can reduce consistency across multiple graders.

Prior to the first meeting, GTAs were tasked with grading an example report using the provided rubric, format guide, and lab manual. These resources were the same materials previous GTAs had access to, allowing researchers to observe grading with minimal instruction. The first training session covered course-specific learning objectives, grading rubric, common lab report issues, and best practices for grading and feedback. GTAs interacted with the material and discussed their evaluations, reflecting on differences in peer grading. Following the first session, GTAs graded a second example report using concepts from the training. These reports were reviewed in the second session, which included a quick review of grading consistency and best practices. This session was shorter and combined with other required training.

2.3 Measurements

Participants were asked to complete three surveys and grade two example reports over the semester. Three timepoints were selected to measure training effectiveness. The first occurred during the week prior to the GTA meeting and established a baseline while collecting demographic information. The second time point occurred during the week following the first training meeting but before GTAs began interacting with students, and measured the immediate impact of the training. The final time point occurred during the final week of the semester, prior to the conclusion of their appointment, to measure the impact of instructional experience gained while serving the GTA role.

The surveys were developed from Smith et al. 2023 article investigating STEM teaching assistant self-efficacy [13] to measure GTA teacher self-efficacy. At each time point, the same survey instrument was administered to assess changes in teacher self-efficacy over time. Participants responded using a scale ranging from not at all confident to very confident. Each survey included the same 14 questions and had one unique free-response question specific to the time point (Appendix). The free-response questions were chosen to gain insight into GTA experience for the course as well as the training.

Example reports were selected by the IOR of the course. Reports were intentionally chosen to include a range of common errors, allowing GTAs to practice identifying errors and providing feedback. Each report was independently graded by the lead investigator and subsequently reviewed by the IOR to establish a control. This control was used to determine which errors would result in point deductions and what level of feedback would be considered helpful to students, consistent with the standards set by the IOR. GTAs independently graded two reports selected by the lead investigator, one before and one after the training. Additional reports were reviewed collectively during the first training session.

From the graded reports, four metrics were calculated. Grades assigned by GTAs to individual rubric sections and the overall report scores were used to measure variance between GTAs as an indicator of grading consistency. The proportion of student errors resulting in point deductions, and helpful feedback on student errors and areas for improvement, was calculated by comparing GTA grading to the control. The number of best practices for grading and feedback used, and the number of unhelpful comments were also counted. These four metrics were used collectively to evaluate GTA grading ability relative to established standards.

2.4 Analysis

2.4.1 Survey Analysis

Survey responses related to teacher self-efficacy were coded into numeric values corresponding to the original range, with not at all confident represented by one and very confident represented by five. A composite score was calculated by averaging each participant's responses, with scores below three considered low, three to four, medium, and above four, high [13]. Pearson correlation was initially considered to compare participants across time points, but inconsistent survey completion limited statistical analysis. Only three participants completed all surveys, so no statistical test was conducted. Instead, results were visualized using a box-and-whisker plot to examine general trends.

Free-response questions were analyzed for common themes; however, responses were similarly inconsistent. Participants were prompted to provide both positive and negative (most and least) responses to each question, though some participants provided only one response. When responses did not explicitly indicate whether they were positive or negative, they were categorized as positive. As a result, these responses were used to inform evaluation of the training experience but were not used to draw formal conclusions.

2.4.2 Grading Analysis

Graded reports completed before and after the training were analyzed to evaluate both grading consistency among GTAs and their ability to grade according to established standards. Completion rates for grading practice were slightly more consistent than the survey participation, as grading was embedded within the training process. However, due to the small sample size and remaining inconsistencies, the Wilcox signed-rank test with a significance value of 0.05 was selected for these measures as well. If the significance threshold was not met, no difference between pre- and post- training performance was concluded.

Grading consistency was evaluated by comparing the scores assigned to each rubric section and the overall report grade. Variance was calculated for each section and the total score for pre- and post-training reports. These variances were then analyzed using the Wilcox signed-rank test with a significance level of 0.05. Moreover, grading ability was evaluated by comparing GTA performance to the control and between pre- and post-training conditions. For the metrics assessing the number of errors identified and the minimum number of helpful comments, the proportion achieved relative to the control was calculated. For the metrics assessing unhelpful comments and use of best practices, raw counts were used. All metrics were analyzed using the Wilcox sign-rank test to determine whether differences existed between pre- and post-training performance.

3. Results

3.1 Teacher Self-Efficacy

Due to the inconsistency of survey completion, statistical significance could not be determined. However, observable trends within the teacher self-efficacy scores can be examined. The average scores for each survey indicate a general increase across the semester (Figure 1). The average teacher self-efficacy score increased the most immediately post-training from 4.01 to 4.22. This upward trend continued through the end of the semester, with an average score of 4.38 at the end of the semester. When accounting for the attrition, there is still an overall increase in scores across the semester; however, the magnitude of increase between timepoints is smaller.

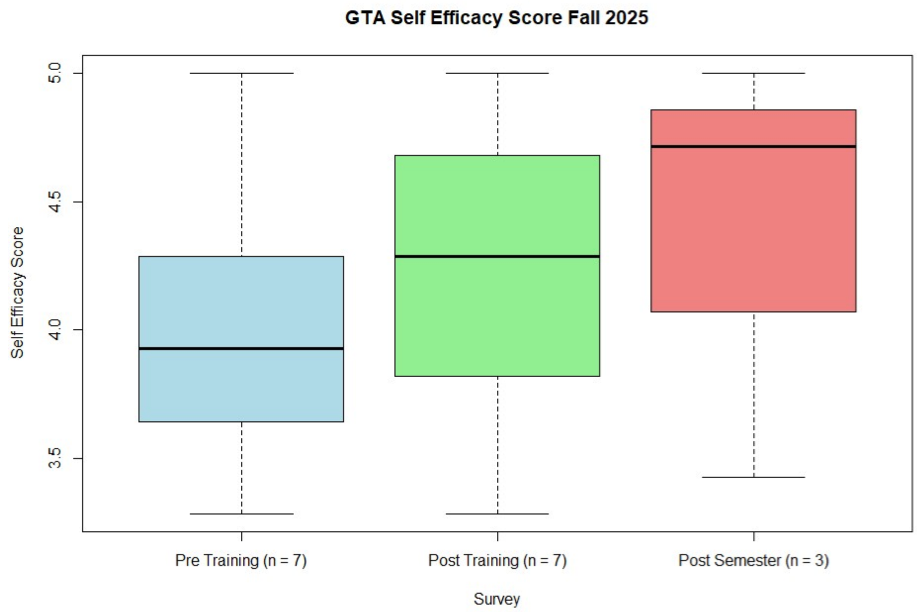

Figure 1: Box and Whisker Plot of teacher self-efficacy scores of Fall 2025 GTAs

Overall, the GTAs in this study demonstrated moderate to high levels of teacher self-efficacy. Three of the seven of the GTAs' teacher self-efficacy scores remained unchanged post-training, indicating that the training did not meaningfully impact teacher self-efficacy for these individuals. Only one GTA's teacher self-efficacy score decreased at the end of the semester. This GTA's score remained stable pre-and post-training at 3.64 but declined to 3.43 by the end of the semester. Among the remaining two GTAs who completed the third survey, one demonstrated no change in teacher self-efficacy, while the other showed an increase.

3.2 Consistency and Ability

Grading consistency was evaluated by calculating the variance between the GTA grades pre- and post-training for each rubric section. The variance between GTAs decreased in all rubric sections except the Figures and References, Language, and Format sections, where the variance increased. The most substantial improvement occurred in the overall score, where the variance decreased from 117.12 to 38.63. Despite this reduction, the Wilcox signed-rank test yielded a p-value of 0.076, indicating no statistically significant difference. Therefore, it cannot be concluded that the training produced a measurable improvement in grading consistency. In addition, when compared to the "correct" score from the control, the median raw score selected by GTAs diverged more after the training.

When analyzing the other four performance metrics, improvements were observed across most measures, with the exception of unhelpful feedback, which showed a slight increase in the number of unhelpful comments. Post-training, GTAs identified approximately 10 percent more errors, demonstrated an 11 percent increase in helpful feedback, and adopted, on average, one additional best practice. While these changes suggest modest improvement, none of the metrics reached statistical significance under the Wilcox signed-rank test. Best practices yielded the lowest p-value at 0.066; however, this value still fails to reject the null hypothesis. Consequently, it cannot be concluded that the training resulted in a statistically significant change in the grading or feedback abilities of the GTAs.

3.3 Evaluation of Training

Three free-response questions were included to capture the GTAs' perceptions of their roles, the training, and their experience across the semester. Prior to training, GTAs were asked what aspects of their role they were most and least looking forward to. Responses were largely consistent, with working as a teaching assistant in a lab course and interacting with students identified as anticipated positive experiences, while grading was unanimously cited as the least anticipated responsibility. Following the training, GTAs were asked to identify the most and least helpful aspects of the training. The most frequently cited beneficial component was the detailed breakdown of the grading rubric, followed by opportunities to compare the grading of example reports. Two GTAs noted that the training was lengthy and included portions that were not engaging. Only one GTA responded to the final free-response question and indicated that the training was particularly helpful for improving their grading and feedback practices.

4. Discussion and Implications

4.1 Self-Efficacy

When examining overall teacher self-efficacy scores, there appears to be a modest increase following the training. However, when the scores are analyzed at the participant level, this improvement becomes less clear. The majority of participants did demonstrate an increase in teacher self-efficacy scores across time points. While the training appears to lower any of the scores, these results do not support the conclusion that the training produced a meaningful impact on teacher self-efficacy. Additionally, among the participants who completed the third survey, the full range of trends was observed, including increases, decreases and no change in scores. As a result, there is no consistent pattern indicating that either the training or subsequent instructional experience positively influenced GTA teacher self-efficacy. Although prior research suggests that professional development can improve teacher self-efficacy [11, 13, 15, 16], the findings of this study do not provide sufficient evidence to support those claims in this context.

4.2 Consistency and Ability

Similar to the findings for teacher self-efficacy, this study cannot conclude that the training resulted in improvements in grading consistency or ability to provide feedback. None of the five evaluated metrics demonstrated a statistically significant difference between pre- and post-training measurements. Of these metrics, variance in assigned grades and use of best practices exhibited the greatest improvement. The substantial reduction in variance for overall report grades, from 117.12 to 38.63, suggests that the training may have contributed to increasing grading consistency. However, this improvement was not uniform across all rubric sections, as two showed an increase in variance. This increase may be attributed to a higher concentration of errors in those sections within the second example report; however, it may also reflect insufficient clarity in the training related to these rubric categories. Moreover, when considering the increase in divergence of raw score, these limitations highlight areas that should be addressed in future iterations of the training to better isolate and evaluate its impact on grading practices.

Use of best practices increased notably following the training. These practices are likely among the easiest elements for GTAs to adopt, as they are explicitly presented and easily referenced. While inclusion of best practices is important for improving grading behavior, they are expected to have the least direct influence on student learning outcomes [20]. The remaining three ability metrics, excluding unhelpful feedback, demonstrated minor improvements post-training. However, these changes were small and did not reach statistical significance. Future iterations of the training should more explicitly address the gap between instruction and implementation to support GTA skill development.

4.3 Training Evaluation

Survey responses indicate that the training was generally well-received by the GTAs. The feedback suggests a need to reduce the overall length of training and increase opportunities for interaction. At the same time, GTAs identified the breakdown of the rubric sections and opportunities to compare grading approaches as particularly valuable components. These elements should be highlighted in future versions of the training. Additionally, results from the consistency and ability analyses suggest that further clarification is needed for Figures and References, Language, and Formatting rubric sections, as well as strategies related to providing constructive feedback. Addressing these areas may improve both consistency and grading ability in future iterations.

4.4 Limitations

This study was performed with a small sample population, and inconsistent participation further limited the strength of the findings. Although some results suggest potential improvement following the training, the lack of consistent participation restricted the conclusions that could be drawn. The methods were selected to ensure GTA participation could be completed within their appointed work hours, limiting data collection methods and preventing qualitative methods such as interviews. While participation in surveys and grading activities was optional, future iterations will revise how the research is introduced to GTAs to encourage more consistent engagement and create opportunities for qualitative methods to improve understanding of GTA experiences and increase feedback on the training. This would also allow investigation of whether teacher self-efficacy influences grading consistency and ability.

Additionally, the example reports were intentionally selected to expose GTAs to a range of errors. However, differences in the distribution of errors between reports may have introduced unintended variability that influenced grading outcomes. In future iterations, example reports selected for individual grading will be selected to include more comparable distributions of errors across rubric sections to reduce this source of variability.

5. Conclusion

Previous research has emphasized the importance of formal training for GTAs; however, this study is unable to confirm a statistically significant impact on teacher self-efficacy, grading consistency, or grading ability. While the results do not support definitive conclusions, trends such as the reduction in grading variance suggest the potential for meaningful improvement.

Future iterations of this work will refine the structure and content of the grading and feedback instruction by expanding opportunities for group discussion and individual practice while shortening the formal presentation to better align with the Cognitive Apprenticeship Model. Also, future iterations will work to better incentivize consistent participation and include more qualitative methods, such as interviews on the GTA experience. Developing effective training for GTAs is critical not only for improving GTA performance but also for enhancing the educational experiences of undergraduate students. Successful implementation of this training within this course may also inform training practices in other courses and departments, supporting broader improvements in engineering education.

References

[1] A. Donadel, "Tenured faculty in steady decline while part-time and graduate workers rise, per report," Apr 2023. [Online]. Available: https://universitybusiness.com/tenured-faculty-in-steady-decline-while-part-time-and-graduate-workers-rise-per-report/

[2] G. Colby, "Data snapshot: Tenure and contingency in us higher education," Mar 2023. [Online]. Available: https://www.aaup.org/academe/issues/spring-2023/data-snapshot-tenure-and-contingency-us-higher-education

[3] A. Moon, H. Jung, F. Marbouti, K. Rodgers, and H. Diefes-Dux, "Undergraduate and graduate teaching assistants' perceptions of their responsibilities - factors that help or hinder," in 2013 IEEE Frontiers in Education Conference (FIE), 2013, pp. 1576–1578.

[4] N. Rao, A. Hosein, and R. Raaper, "Doctoral students navigating the borderlands of academic teaching in an era of precarity," Teaching in Higher Education, vol. 26, no. 3, pp. 454–470, 2021. [Online]. Available: https://doi.org/10.1080/13562517.2021.1892058

[5] E. E. Schussler, Q. Read, G. Marbach-Ad, K. Miller, and M. Ferzli, "Preparing biology graduate teaching assistants for their roles as instructors: An assessment of institutional approaches," CBE—Life Sciences Education, vol. 14, no. 3, Sep 2015.

[6] M. K. Anderson, R. J. Anderson, L. S. Tenenbaum, E. D. Kuehn, H. K. Brown, S. B. Ramadorai, and D. L. Yourick, "The benefits of a near-peer mentoring experience on stem persistence in education and careers: A 2004-2015 study," The Journal of STEM Outreach, vol. 1, no. 1, Feb 2019.

[7] J. B. Dillard, K. Sadek, and K. Muenks, "Undergraduate perceptions of graduate teaching assistants: competence, relatedness, and autonomy in practice," Higher Education Research & Development, vol. 43, no. 1, pp. 32–47, 2024. [Online]. Available: https://doi.org/10.1080/07294360.2023.2215169

[8] G. Ramasubramanian, "Development and evaluation of a professional training module for chemistry graduate teaching assistants," Ph.D. dissertation, University of Oregon, 2022.

[9] A. Bandura, Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[10] S. Gibson and M. H. Dembo, “Teacher efficacy: A construct validation.” Journal of Educational Psychology, vol. 76, no. 4, p. 569–582, Aug 1984.

[11] M. R. Connolly, Y.-G. Lee, and J. N. Savoy, “The effects of doctoral teaching development on early-career stem scholars’ college teaching self-efficacy,” CBE—Life Sciences Education, vol. 17, no. 1, p. ar14, 2018, pMID: 29440074. [Online]. Available: https://doi.org/10.1187/cbe.17-02-0039

[12] D. B. Morris and E. L. Usher, “Developing teaching self-efficacy in research institutions: A study of award-winning professors,” Contemporary Educational Psychology, vol. 36, no. 3, pp. 232–245, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0361476X10000561

[13] C. R. Smith, D. Menon, A. Wierzbicki, and J. M. Dauer, “Exploring stem teaching assistants’ self-efficacy and its relation to approaches to teaching,” CBE—Life Sciences Education, vol. 22, no. 1, p. ar6, 2023, pMID: 36637378. [Online]. Available: https://doi.org/10.1187/cbe.22-06-0115

[14] J. Wright and D. Summers, Lauren N.and Wilson, “It’s not just what ta’s know: Exploring the role of teacher efficacy among engineering ta’s,” in 2019 ASEE Annual Conference & Exposition, no. 10.18260/1-2–33032. Tampa, Florida: ASEE Conferences, June 2019, https://peer.asee.org/33032.

[15] P. H. P. Chiu and P. Corrigan, “A study of graduate teaching assistants’ self-efficacy in teaching: Fits and starts in the first triennium of teaching,” Cogent Education, vol. 6, no. 1, p. 1579964, 2019. [Online]. Available: https://doi.org/10.1080/2331186X.2019.1579964

[16] S. E. DeChenne, N. Koziol, M. Needham, and L. Enochs, “Modeling sources of teaching self-efficacy for science, technology, engineering, and mathematics graduate teaching assistants,” CBE—Life Sciences Education, vol. 14, no. 3, p. ar32, 2015, pMID: 26250562. [Online]. Available: https://doi.org/10.1187/cbe.14-09-0153

[17] V. Dennen and K. Burner, “The cognitive apprenticeship model in educational practice,” Handbook of research on educational communications and technology, pp. 425–439, 01 2007.

[18] S. Manuel-Dupont, “Writing-across-the-curriculum in an engineering program,” Journal of Engineering Education, vol. 85, no. 1, pp. 35–40, 1996. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.1996.tb00205.x

[19] D. Rus, “Developing technical writing skills to engineering students,” Procedia Technology, vol. 19, pp. 1109–1114, 2015, 8th International Conference Interdisciplinarity in Engineering, INTER-ENG 2014, 9-10 October 2014, Tirgu Mures, Romania. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212017315001590

[20] J. T. Jørgensen, B. Gammelgaard, and F. V. Christiansen, “Teacher intentions vs student perception of feedback on laboratory reports,” Journal of Chemical Education, vol. 100, no. 10, pp. 3764–3773, 2023. [Online]. Available: https://doi.org/10.1021/acs.jchemed.2c01148

[21] Z. Yan, C. W. Lum, R. T. L. Lui, S. S. Chu, and M. Lui, “Measuring teaching assistants’ efficacy using the rasch model,” Journal of Applied Measurements, 2015.

Appendix: Survey Questions

Self-Efficacy Questions[13, 21]:
Scale:

1. Not at all confident
2. Slightly confident
3. Neutral
4. Moderately Confident
5. Very Confident

How confident am I in my ability to:

1. Think of my students as active learners, which is to say knowledge builders rather than information receivers
2. Actively engage my students in the learning activities that are included in the teaching plan/syllabus
3. Promote a positive attitude towards learning in my students
4. Provide support/encouragement to students who are having difficulty learning
5. Encourage the students to interact with each other
6. Show my students respect through my actions
7. Let students take initiative for their own learning
8. Appropriately grade my students' lab reports
9. Provide constructive feedback on my students' lab reports
10. Evaluate accurately my students' academic capabilities
11. Clearly identify the course objectives
12. Provide my students with detailed feedback about their academic progress
13. I can provide an alternate explanation or example when students are confused
14. Stay current in my knowledge of the subject I am teaching

Free Response:
Survey 1 – Pre Training
What are you most and least looking forward to in your Teaching Assistant position?

Survey 2 – Post Training
What was the most and least helpful during the training session?

Survey 3 – End of Semester
What was the most helpful this semester in improving your ability to grade and give feedback?

Demographic Questions:
What is the highest degree that you have completed?

- Bachelors
- Masters
- Other

What is your current degree?

- Masters
- PhD
- Other

How many semesters have you been a teaching assistant?

- 0 (First Semester)

- 1
- 2
- 3
- 4 +

Did you request this teaching assistant position?

- Yes
- No

After the completion of your current degree do you plan to ?

- Continue my education
- Go into industry
- Go into academia
- Other