

3 Module 3: Teaching Assistants as Graders

3.1 Summary

This document sheds light on the issue of employing teaching assistants as graders without providing adequate training. We have detailed the problem by referencing relevant literature and presenting specific cases we have encountered. Furthermore, we have reviewed current rules and regulations and compiled suggestions to help the program management formulate a comprehensive policy or set of best practices. Additionally, we have offered advice for examiners on how to effectively utilize teaching assistants in grading. Lastly, we introduce a method to detect potential grading inconsistencies.

3.2 Introduction

The computer science program has seen massive growth in recent years. The change posed a logistical challenge for the examiners and called for hiring more lecturers. To provide education at scale, support from teaching assistants (TAs) became essential. In the current academic year, the TAs reported almost 30,000 working hours. Teaching assistants are hired for all quarters, but a majority of them are hired to help with the first-year courses since those have a higher number of students, and also, the courses are not very advanced.

3.2.1 Teaching Assistants and their Roles

The population groups of teaching assistants are diverse in their education levels. The following is a categorization of teaching assistants:

- Undergraduate students(UTA): These are usually year two or year three students. In most cases, they have followed the same courses they are assisting.
- Graduate(Master) students (GTA): These are master's students. Usually, they are former UT students who have followed the same course (but it is not always the case).
- Ph.D. Candidate (Ph.D.): Pursuing their Ph.D. and may not have followed the same courses they are assisting

Ph.D. students are not considered as teaching assistants at UT, rather they are referred to as lecturers, tutor, or instructors.

Teaching assistants have various roles depending on the context in which they were hired. Some of their roles include tutoring, assisting the teacher in the classroom, leading small student groups, preparing assignments, auditing assignment descriptions, **grading assignments, exams and projects**, and providing general administrative support¹.

In this memorandum, we outline the challenges and offer recommendations for various stakeholders regarding the deployment of teaching assistants to grade assignments, exams, and projects. The memorandum is especially directed towards the stakeholders from the computer science program but most recommendations are generally applicable.

3.3 TA as Graders-Problem Background

Due to the large size of the required teaching assistants (in the first year modules, on average, 70-100 teaching assistants are hired), it is difficult to determine the understanding level of the subject. Therefore, the program management have introduced rules to hire TA. Besides the rules governed by the labor law TCS program follows the following rules:

¹ Trevor Kerry. 2005. Towards a typology for conceptualizing the roles of teaching assistants. Educational Review 57, 3 (2005), 373–384.

- Teaching assistants (TAs) to be hired should have completed the course or module for which they are being considered. Typically, exceptions are made for Master's students who have completed a similar course at a different university or whose subject expertise has been verified through alternative means.
- When considering second-year students to TA first-year courses, they must have successfully completed the first year, including math. Exceptions can be made by the program director under special circumstances.

Teaching assistants are helping with grading. In most cases they help grading assignments and practical works(generally termed sign-offs). However, they also help grade projects and written exams.

Wald and Harland² have categorized assessments into two types when it comes to deploying teaching assistants as a grader a) Assessing factual information that requires recall b) Assessing items that require interpretative judgments

If the correct answer is available the first one is easy to grade. The second one is problematic since there might not be single correct answer and the correctness of an answer depends on interpretations. Some pointers to correct answers are useful to help the teaching assistants. However, the teaching assistants (and also the lecturer) must have very good understanding of the subject to provide a interpretive judgement translatable to grade or score.

3.3.1 TA as Graders According to EER

The following is the part of rules and guidelines to deploy TA as grades:

1. In some situations, teaching assistants may be deployed as graders.
2. The following conditions must be met:
 - a. the student assistant has, to the examiner's judgment, a good level of understanding of the subjects underlying the tests;
 - b. the examiner has drawn up detailed instructions for grading the tests;
 - c. the examiner has personally assessed several assignments in advance and discussed these with the student assistant(s);
 - d. marks of 5.0 and 6.0 (or inbetween) must be reviewed by the examiner;
 - e. once the marking is complete, the examiner will carry out a random check to ensure the marking is appropriate;
 - f. a student assistant will inform the examiner of any conflicts of interest and will exempt him/herself from grading the test in question.
 - g. the examiner takes steps to ensure the careful handling of the material to be assessed by the student assistant(s).

It is apparent from the above that the examiner is the judge to determine the level of understanding for teaching assistance. However, as a normal practice TAs are selected based on their availability on the basis of the assumption that they must possess the adequate knowledge of the subject.

The introduction of the new contract system has added further complexities. It allocates a fixed number of weekly hours for a TA, creating challenges when trying to find a TA with available hours left in their contract. It's important to note that results must be announced within 10 working days. This tight 10-day timeframe, coupled with the limitations on TA working hours, often results in teachers having to collaborate with whichever TA is available for grading.

² Wald, N., & Harland, T. (2020). Rethinking the teaching roles and assessment responsibilities of student teaching assistants. *Journal of Further and Higher Education*, 44(1), 43-53.

Majority teaching assistants are UTA (almost 60-80%). The practice is against the one reported in literature, given that the major TA population is undergraduate students. Hogan and Norcross³ argue that the UTA must be deployed as a grader with extreme caution and under strict supervision. They must not provide summative feedback and grade only the anonymized assignments. The practice is not followed at all at UT since it is challenging to hire GTAs.

3.3.2 Impacts on the quality of assessment

The following can be impact on the quality of assessment for using teaching assistants for grading:

- **Expertise:** Teaching assistants may not have the same level of expertise as full-time faculty members in their subject area, which could affect the accuracy and thoroughness of their assessments.
- **Consistency:** Teaching assistants may have different teaching styles and expectations, which could lead to inconsistency in assessment practices.
- **Time and resources:** Teaching assistants may have less time and resources to devote to assessment, which could affect the quality and depth of their evaluations.
- **Training and support:** It is important that teaching assistants receive adequate training and support in assessment practices in order to ensure the quality of their evaluations.

In the appendix, we provide examples to underscore the impact observable by teachers and examiners engaged in multiple modules.

3.3.3 Impact in terms of Quality of assessment Pyramid

Transparency, reliability, and validity are three important factors that can affect the quality of assessment in higher education. When teaching assistants are used as assessors, these factors may be impacted in the following ways:

- **Transparency:** If teaching assistants are used as assessors, it is important that they are clear about their expectations and that the assessment process is transparent to students. This could be achieved through providing clear grading rubrics, giving feedback on assignments, and being available to answer questions about the assessment process.
- **Reliability:** The reliability of an assessment refers to the consistency of the results it produces. If teaching assistants are used as assessors, it is important that they are consistent in their evaluation practices to ensure the reliability of the assessment. This could be achieved through providing clear guidelines for assessment, using the same criteria for all students, and using multiple sources of evidence to make judgments about student performance.
- **Validity:** The validity of an assessment refers to its ability to accurately measure what it is intended to measure. If teaching assistants are used as assessors, it is important that they are well-trained and have the expertise necessary to accurately assess student performance in their subject area. This could be achieved through providing adequate training and support in assessment practices, ensuring that teaching assistants have the necessary subject-specific expertise, and using multiple sources of evidence to make judgments about student performance.

3.3.4 Roles and Responsibilities

The role of different actors such as the program management, examination board, support staff, and teaching assistants themselves can be important in the deployment of teaching assistants for assessment in higher education. Here are a few potential roles for these actors:

- **Program management:** The program management team is responsible for overseeing the overall quality and effectiveness of the academic program. They should take decisions about the use of teaching assistants as assessors, including setting policies and procedures for assessment and providing support and resources for teaching assistants.

³ Hogan, T. P., & Norcross, J. C. (2012). *Undergraduates as Teaching Assistants. Effective college and university teaching: Strategies and tactics for the new professoriate*, 197.

- **Examination board:** The examination board is responsible for overseeing the assessment process and ensuring that it is fair and consistent. To ensure this, *they are responsible for appointing examiners*. Thus, they must be involved in decisions about the use of teaching assistants as assessors, including advising on policies and procedures for assessment and providing support and resources for teaching assistants.
- **Support staff:** Support staff may play a variety of roles in the assessment process, including providing administrative support, managing assessment data, and assisting with the development and implementation of assessment policies and procedures.
- **Teaching assistants:** Teaching assistants who are involved in assessment have a number of responsibilities, including understanding and following assessment policies and procedures, providing clear feedback to students, and ensuring the reliability and validity of their evaluations.
- **Examiner:** The examiner plays a crucial role as the EER delegates the hiring of TAs to the examiner's discretion. Additionally, they are the primary overseers of the process and bear the responsibility of ensuring fair and consistent grading.

3.4 Advise for Program Management

Examiners typically employ TAs with careful consideration, equipping them with detailed rubrics to ensure a fair and consistent grading process. Despite these precautions, issues still arise in the grading process, as illustrated by the examples provided in the Appendix. We have organized our recommendations, primarily directed towards program management, into four main categories.

- Deploying TA for grading-hiring practice
- Providing appropriate training to TAs
- Facilitating examiners to detect potential problem
- Anonymization and Grading Session

We will discuss all four in the coming sections motivating their occurrence and providing good practices from literature.

3.4.1 Deploying TA for grading-hiring practice

EER has established rules for using TAs in grading roles. However, there is currently no system in place to ensure these rules are consistently followed. While the examination board has a defined process for appointing regular staff as examiners, the guidelines become less clear when it comes to appointing TAs. In these instances, the responsibility often shifts to the individual examiner without clear direction. Here are a few recommendations for the program management to establish a clear policy on basis of these recommendations:

- Prioritize hiring graduate students for grading exams.
- When employing an undergraduate student as a grader, ensure the student's grade is within the top 10% of their cohort for the subject they will be grading.
- Implement a monitoring mechanism by preparing a report on the TA hiring process. This report could be incorporated into the examination log and uploaded to the Seshat system.

Recommendation 1 and 2 are based on recently published paper on the topic⁴.

Additionally, In the literature, the prevalent recommendation is for two teaching assistants (TAs) to independently grade a student's work. If the grades provided by both TAs are similar, the student receives the average of the two grades. If there is a significant discrepancy between the grades, a third evaluator should review the work. While this approach might increase costs due to additional TA hours, the enhanced reliability in grading could justify the expense.

⁴ Wald, N., & Harland, T. (2020). Rethinking the teaching roles and assessment responsibilities of student teaching assistants. *Journal of Further and Higher Education*, 44(1), 43-53.

Note: Remindo offers a feature that allows two TAs to independently grade a student's assignment. However, this feature is not well-documented, and we stumbled upon it by chance.

3.4.2 Providing appropriate training to TAs

The examiners are expected to attain training for creating and administering exams as part of their UTQ. In contrast, there are no such requirements for teaching assistants. It can be argued that the examiner is solely responsible for creating and administering exams, and TAs help them to grade these exams. A well-balanced exam that is appropriately administered could be invalid since the grading was not reliable. The program management can design general training that emphasizes the role of assessment in education. The training should also include components related to giving feedback, dealing with plagiarism, and accommodating student concerns.

Note: The current program director has suggested specialized training for the varied roles of TAs, as illustrated in Figure 1. The initial training focuses on the 'Facilitator TA' role, and there's also specific training being designed for 'Grader TAs'. The comprehensive training materials and course outlines have not been released yet.

Assistant § 3.1	Expert	§ 3.2	answer subject matter questions
	Grader	§ 3.3	perform summative assessment
	Facilitator	§ 3.4	provide normative assessment
	Analyst	§ 3.5	explore options and improve processes
	Shaper	§ 3.6	brainstorm, be creative and critical
	Coach	§ 3.7	get to the root of the problem

Figure 1: Different roles and associated trainings for teaching assistants⁵

3.4.3 Facilitating examiners to detect potential problem

The computer science program is using Remindo systems for their most digital exams. The system gathers a large amount of data that can provide various insights concerning grades and the grading process.

The digital examination system **Remindo** provides several statistics related to the exam. However, no statistics could be computed concerning the grader. A new feature that allowed the examiner to review the **grading history of an examination** item(for example, who has graded and what are the changes to the grade). The feature is restrictive because the changes can be viewed only on a per question and per student basis. It would be helpful to export such a log for detailed analysis to compute, inter and intra-ratter variation in grading. The analysis could also help the *examination board* audit changes in the grading and see whether these changes are justified or not.

To enhance grading consistency, the program management should request additional features in Remindo. The following is a non-exhaustive list of proposed features that could aid examiners:

- Percentage of exams graded by an individual teacher.
- Number of personnel grading a specific question.
- Visualization tools for grading variation, such as bar graphs (see Appendix for examples).
- Computation of statistical significance parameters to assess variation in grading across different personnel.
- Track grading time per question to identify potentially problematic questions that take a longer time to grade.

⁵ Vadim Zaytsev, Roles and Competences of Teaching Assistants

- Comparison of grading trends over time to identify any deviations or inconsistencies in grading patterns.
- Feedback mechanism for TAs to communicate uncertainties or ambiguities they encounter while grading.

Not only do these features support the examiners, but they also empower the examination board to effectively audit the grading process.

3.4.4 Anonymization and Grading Sessions

The student's(to be graded) information must be anonymized. In the digital exam, **anonymization** is trivial. However, the practice is not followed strictly. In consultation with the examination board, the program management formulates a procedure to anonymise the non-digital exams and project work.

Encourage grading sessions. It is becoming common knowledge among examiners that **grading sessions** are more effective and efficient. The active discussion during the grading session reduces the ambiguities and increases the uniformity of the grading.

3.5 Advise for the Examiners

In the absence of formal training the following should be ensured by the examiner:

- Rubric and Model Answer Development: Construct a detailed rubric with a corresponding model answer. Train teaching assistants on how to use these rubrics effectively. If a student's answer does not align with the rubric but has merit, TAs should discuss it with the examiner.
- Binary Grading Criteria: It might be beneficial to establish grading criteria that are binary. This approach reduces the potential for subjective judgments by teaching assistants.
- Preliminary Review of Answers: Before finalizing the grading criteria, examine a selection of students' answers. Doing so will help recognize answer variations and refine the grading criteria.[This is already required by EER]
- Feedback Examples: Offer examples of feedback pertinent to the questions that will be graded. This ensures teaching assistants provide consistent and clear feedback.
- Student Review Sessions: Promote student attendance at review sessions. This not only aids students in understanding their performance but also helps identify any grading inconsistencies.
- Grading Sessions: Organize dedicated grading sessions where teaching assistants come together in a designated room, rather than grading individually at their leisure. Promote active discussions among TAs and examiners during these grading sessions.

While the aforementioned recommendations are primarily for examiners, the **program management** can integrate them into their policies for utilizing teaching assistants as graders.

Appendix

A Method to analyze Exam grades

Our analysis of the exam grades is focused on capturing variations in the grades given by a TA. We focus on capturing inconsistencies in grades assigned by different TAs. Also, we investigate the differences in the grading patterns between UTA and GTA. Some TAs have been working for the program for a longer time. They were UTAs, and are now GTAs. We applied the analysis to two exams.

Exam A

The BCS exam was graded by 22 TAs and three teachers. Out of 22 TAs 15 were UTA while 7 were GTA. After that, the questions were divided among TAs based on their preferences. At least two TAs graded each question (to emphasize, the set of students each TA was grading was different). Some questions had

more tasks and required more verbose answers than others. Consequently, more TAs are needed to grade them. As soon as a TA is done grading a question and still willing to do more grading, a new question is assigned to them.

Since we wanted to capture variation in grading, we adopted the box plot. Box plots are handy for visualizing variations. We created one box plot for each TA and combined a box plot for the same questions in one figure. TAs names are removed to maintain privacy. For the questions that are graded by two TAs, we have also applied a two-sample t-test to find out whether the difference in mean grading is statistically significant or not. Note that the test is only suitable for pairwise comparison. Therefore, we have applied it only to questions that, at most, two teaching assistants grade.

The results of the tests show that the difference in mean grades is not statistically significant. We conclude that the TAs graded questions consistently. However, when a question is graded by more than two TAs, the variation is higher. Since we cannot judge significance using a t-test, we opted to compare it with the total grade. The implicit assumption is that if, on average, a TAs is giving a lower (or higher grade respectively) while the total grade has an opposite trend, then the TA might be grading harshly (or generously respectively). We emphasize that before drawing any conclusions, a sample must be checked by the teachers to verify the claim.

To illustrate, we provide one example. Figure 2 show a box graph for one question since the question has only 6 points while the total score is 10. To compare both graphs, we scaled both between 0 and 1.

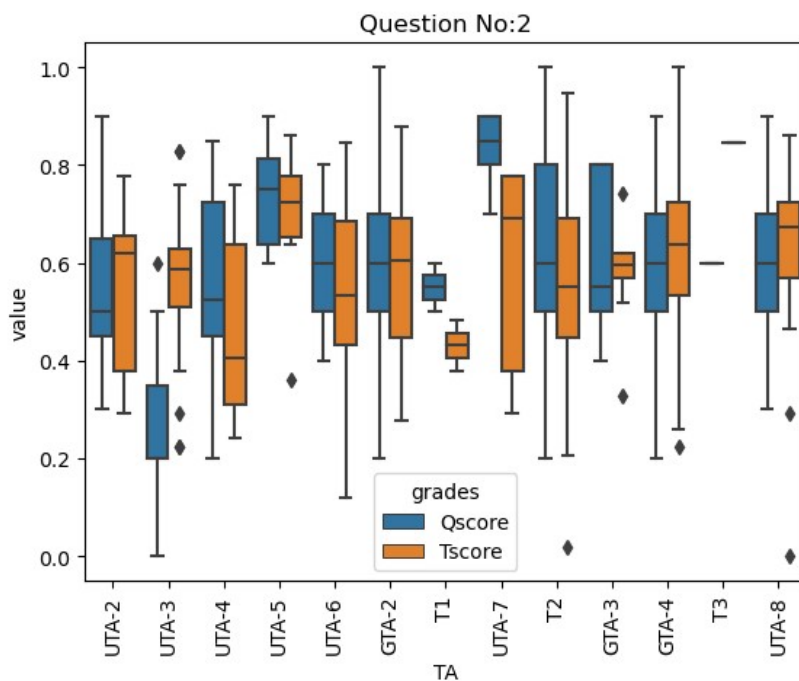


Figure 2: The plot shows grades obtained by student for the question (blue box plot) and the total score obtained by a student (orange box plot). Both scores are scaled between 0 and 1.

As can be seen from Figure 2 UTA4 have assigned lower grade for this question, while for the same student population, their overall score is much higher. Thus, this grade section requires further investigation either by a teacher or another TA. Another interesting observation is for T3, which represents a teacher with only one data point. It might be the case that this question is borderline, and the teacher decided to review and regrade.

We also like to investigate the claim that GTAs are better in grading exams since they have higher qualification. We were not able to identify any significant difference for the questions shown in Figure 2.

Exam B

The BIT exam represents a different style of use of TAs. For this exam, only one GTA was used, who was a previous undergraduate student. The teacher has a more personal approach to hiring TA and relies on a small team of trustworthy students. Since the number of students lies between 55-120 for this course, the grading work is significantly smaller than for the BCS course. Here we analyze the exams for three years to capture the grading practice for a TA.

For most of these exams, there is no significant difference in grading from different TAs. Figure 3 shows a box plot for a question that appeared in the 2021 exam. The question is related to the activity diagram. The question is accompanied by a detailed rubric that converts an open-ended grading task into a binary grading.

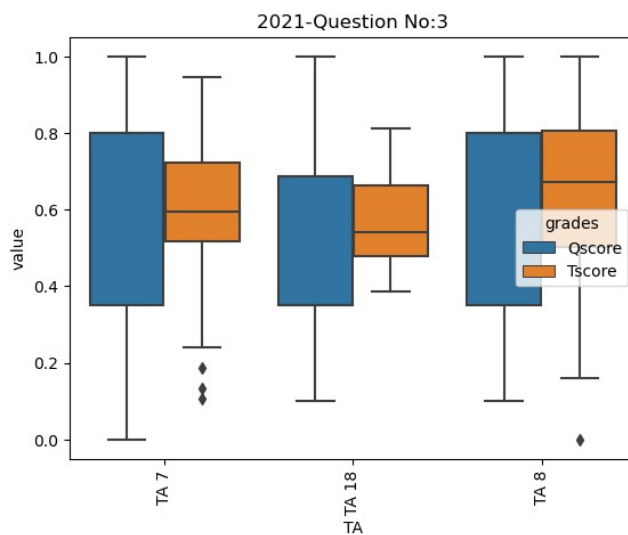


Figure 3: (For exam B) The plot shows grades obtained by student for a question (blue box plot) and the total score obtained by a student (orange box plot). Both scores are scaled between 0 and 1.

Besides the clear rubric, there is a difference in grading from different TAs. TA18 give lower grades, while TA7 have higher variation in their grading.

During our exams, we have consistently engaged certain teaching assistants for grading. From this practice, we have observed that senior TAs, those who have graded the same type of exam multiple times, are better equipped to evaluate answers not specifically addressed in the provided rubric. However, they tend to make errors when grading open-ended questions, as they sometimes miss the various forms that mistakes can manifest in. On the other hand, junior TAs often adhere strictly to the given rubric. This strict adherence can cause them to overlook or not adequately reward partially correct answers.