

Training Teaching Assistants in Assessment of Lab Assignments in Computer Science at a Swedish University

Emma Riese, and Viggo Kann

KTH Royal Institute of Technology

ABSTRACT: Teaching assistants (TAs), students who assist the faculty, are widely used in computer science (CS) courses. Previous studies have, however, shown that TAs could be poorly prepared and need training. Particularly, we have previously found that one of the areas where our TAs experience uncertainty is when assessing students' oral presentations of their lab assignments. Based on that result and by interviewing course coordinators, we have developed and offered TA training workshops about assessment in CS. We invited our TAs in the introductory CS courses to participate voluntarily. By distributing pre-workshop surveys at the beginning of each semester, and post-workshop surveys at the end of the semesters, to both workshop attendees (50) and non-attendees (44), we studied how the TAs conducted the assessments and what impact the training workshop had on their self-reported practice. Pre- and post-workshop surveys had 11 identical statements that the TAs were asked to rate on a 7-point Likert scale. We also conducted interviews with four workshop attendees and three non-attendees. The results showed a significant difference between the two groups in the post-workshop survey: workshop attendees disagreed more with the statement "I try to assess students' understanding rather than the program", which is more in line with the instructions given. In addition, when comparing pre-and post-workshop answers, the workshop attendees stated that they were less inclined to ask for help, experienced that the lab instructions were not detailed enough, and were more inclined to ask questions that convinced them that the students had written the program themselves. In the control group, no significant differences between pre-and post-tests were found.

1 INTRODUCTION

Teaching assistants (TAs) are widely used in computer science (CS) courses, especially in courses that are given to a large number of students [14, 18, 22]. TAs are typically undergraduate or graduate students who have previously enrolled in the same or a similar course, working parallel with their own studies [21]. The TAs' work tasks could differ between universities and courses, but could include conducting tutorials, tutoring students, and grading assignments [18, 21, 28, 32]. Even though the TAs are responsible for many of the interactions with students, previous studies have shown that the TAs could be poorly prepared and offering TA training has been suggested as a way to assist the TAs with their professional development [13, 23, 32]. Some institutions have also reported on initiatives of TA training [6, 14, 17, 19], but this is an area that is not fully explored and researched. Few systematic evaluations of TA training have been carried out.

The university where this study was conducted offers a variety of engineering programs. TAs are used in almost all CS courses at the undergraduate level and in some more advanced courses. One of the most typical work tasks for our TAs is to assess students' oral presentations of their lab assignments. The TAs are expected to both give the students feedback and grade their assignments. A previous interview study with eight TAs from our introductory courses revealed that the TAs experienced some uncertainty about these assessments and work in the lab session, for instance, that the grading instructions lacked necessary details, how to detect and handle plagiarism, time pressure, and that failing a student is difficult [29]. Based on that result, we decided to develop and offer a training workshop about assessment in CS for all TAs teaching in our programming courses (given in Swedish) during the academic year 2018/2019. We chose the introductory courses as our primary focus because many of our TAs start their TA careers there. Since the previous study used a qualitative approach with a small sample size [29], we were also

interested in investigating how generalizable the experienced uncertainties and practices were. This paper describes the intervention and our findings from our data collections, addressing the research questions (RQs):

1. How do TAs conduct and handle the assessments of students' lab assignments in introductory CS courses within a Swedish university?
2. What material and instructions should be provided by the course coordinators for the TAs to be able to assess lab assignments?
3. To what extent can a training workshop change the TAs' practices and experiences regarding the assessment of CS lab assignments?

The results of this study could guide future initiatives and programs of TA training and give insight into important considerations when planning to use TAs to conduct assessments of lab assignments in CS.

2 RELATED RESEARCH AND THEORY

Programs for employing students as TAs have been reported as successful at multiple institutions [6, 10, 21]. It has been argued that TAs make it more comfortable for the students since students can experience it as easier to ask other students for help rather than senior professors [9, 28]. The TAs are typically closer in academic age to the students and were themselves learners in the subject more recently [28], and TAs have also been shown to view themselves as both a friend and a teacher to their students [30, 32]. This could, however, lead to conflicts of interest arising when it comes to grading [30, 32]. Previous research has also shown that TAs have been poorly prepared for their role and have not received sufficient training [8, 23, 29], despite the fact that TAs themselves also perceive training as a factor that would aid their professional development [20]. Within higher education, formal pedagogical training has also been shown to positively affect teaching practices among faculty members [26]. TAs' development has also been studied, and formal mentorships have been suggested as training [25]. Social aspects have also been found to impact the job satisfaction of TAs in a computer lab setting [24]. To recruit students with sufficient knowledge and competency to be TAs is, of course, desirable and a suggestion on a rubric to do so has been presented in [18]. Even though some institutions report on a high interest among the students to become TAs [33, 36], it is not realistic to recruit only TAs who are already skilled and experienced teachers. It has been suggested that TA training can further prepare the TAs for their new role and make them develop the skills and competencies needed [13]. Initiatives of TA training have also been reported [6, 7, 13, 14, 17, 19] where an introduction day, seminar, or course has been offered. Another approach to train TAs and make them reflect upon their teaching practices is to use a reflecting diary [35]. This approach has been used for TAs who also had teaching experience and as part of a more extensive TA training course [35]. To allow the TAs to work collaboratively has not only shown beneficial to job satisfaction [24] but also from a reliability aspect, where assessments conducted by TAs in a group setting were compared to assessments that were done individually [15]. To the authors' knowledge, no prior initiatives of TA training specifically about assessment of CS lab assignments have been reported, a gap we aim to fill with this paper. Assessment has, however, been found to be particularly challenging for TAs [29, 32].

In this paper, we use a broad definition of assessment, which includes formative assessment - feedback to help the students in their learning process, and summative assessment - grading of how well students fulfilled the learning outcomes [11]. When grading, an analytic approach, using grading criteria and grading rubrics, is used quite commonly in higher education [34]. A literature review [27] on the usage of grading rubrics in higher education showed that it is believed that grading rubrics can make it easier to keep the assessment objective. However, some instructors remain skeptical and unwilling to use them [27]. To use a language that is clear to both the students and the graders is highlighted to be one of the challenges when designing grading rubrics since ambiguous phrasing can open for different interpretations [27]. The assessments throughout a course should also be aligned with the learning objectives and the learning activities, which is referred to as constructive alignment [2]. When assessing, it is also crucial for all parties to know what is being assessed, which skill or knowledge, and how that is, or at least can be, shown to the grader. One way to categorize this is by using the levels in Bloom's taxonomy [3] or Bloom's revised taxonomy [1]. In short, Bloom's revised taxonomy consists of two dimensions: the cognitive process dimension and the knowledge dimension. The cognitive process

dimension contains six categories: *remember*, *understand*, *apply*, *analyze*, *evaluate*, and *create*, in that particular order, where each category is more cognitively complex than the prior one. The knowledge dimension consists of four categories: *factual*, *conceptual*, *procedural*, and *metacognitive*, going from concrete to abstract [1]. When the students are developing programs to solve the assignments, the categories *understand* and *create* from the cognitive process dimension are the most apparent ones and what the assessment is typically focused on. In the knowledge dimension, the assignments can span through all categories, much depending on the course and assignment at hand, but factual, conceptual, and procedural are the most apparent ones in the studied setting.

3 RESEARCH SETTING

This section describes the setting in which the study was conducted, the timeline for the project, and the content of the workshop.

3.1 Course Structures and the Roles of the TAs

Our university offers a number of five-year engineering programs. In most of these, there is a mandatory programming (CS1) course during their first year, and an elective CS2 course in the second or third year. CS-majors, however, enroll in four CS courses already in their first year, followed by more advanced courses. Each course has a course coordinator who plans the course, including all examinations, and gives a series of lectures. Each week there is also a tutorial (one or two hours) and a two-hour computer lab session. In the tutorials, the students are divided into smaller groups of between 12 to 40 students, led by a TA or the course coordinator. The tutorials are typically problem-solving sessions where the introduced concepts are put into practice. In the computer lab sessions - the main focus of this initiative - the students are working, usually in pairs, on lab assignments while TAs are available for help and guidance. Typically, a digital queue system is used. The students would usually also present their solutions orally to a TA. In some lab assignments, the students' programs are automatically tested by an online judge [12]. Otherwise, the TAs are also responsible for checking the programs' functionality. The CS1 courses are concluded by a larger individual programming project that is also peer-assessed. The projects are assessed on a grading scale A-F; however, a majority of the lab assignments in the courses are graded on the scale pass/fail. The TAs in these courses are typically undergraduate students or Master's students and, in a few cases, doctoral students (during the academic year 2018/2019, three of the 94 TAs). The course coordinators typically recruit the TAs. Currently, there is no standard matrix used for the recruitments, such as in [18]. Each course coordinator is instead free to decide which qualities and skills they deem necessary for their TAs to have and hire those who they see fit.

3.2 Timeline of Project

The timeline for the project is described in Table 1 below.

Table 1. Timeline of project

Month	Task(s)
July - August 2018	Interviews with three of the course coordinators regarding their expectations of their TAs. Development of the training workshop, based on the described needs (identified in Riese [29]), and the conducted interviews with the course coordinators.
August - September	First offerings of the workshop. The pre-workshop survey was distributed to the workshop participants (before the workshop started) and digitally distributed to all the TAs who were invited to the workshop but did not attend.
October	Post-workshop survey to all TAs who answered the pre-workshop survey.
November - December	Workshop with course coordinators/instructors (6 participants). All the workshop material was presented, and the course coordinators did all exercises. They were also asked to give us feedback on missing aspects or possible improvements. The feedback was solely positive, confirming the need for the workshop, its content, and its structure.
January - February 2019	Workshop given again to TAs who were scheduled to work during the spring semester. New pre-workshop survey was distributed.

Mars - July 2019	Interviews with TAs, post-workshop survey distributed to all TAs who answered the pre-workshop survey in the latest iteration.
------------------	--

3.3 The Workshop

The workshop¹ was given on campus as a two times 45 minutes session, twice at the beginning of each semester, and was also given once to the course coordinators to validate the workshop content. The number of participants was between 4 and 29. The workshops consisted of three parts. In the first part, we discussed why we assess, different goals with assessments and different types of assessment. The concepts formative and summative assessment [11], validity, reliability, peer, and self-assessment were presented, and some aspects of each concept were discussed. One example of such an exercise is the following, in which the TAs were first asked to discuss in smaller groups and then asked to share their answers:

How can we assess...

...that the program is working according to the specification?

...that the students know how the program works?

...that the students have written the program themselves?

...that students can explain relevant concepts?

This exercise highlights that what we want to assess impacts how the assessment should be carried out. It was also done to assist the TAs in distinguishing between these aspects, which the course coordinators workshop participants validated as typical in our labs. A previous misconception has been that if the students can explain the code, they must also have written it [29] (points 2 and 3 above).

The second part of the workshop consisted of reflection and discussion tasks around the following topics:

- How to detect and handle plagiarism.
- How automatic testing of the students' code [12] and peer-review could influence the assessment.
- How to assess students who have worked together.
- How to give oral and written feedback.
- How to handle time pressure in assessment situations.
- What to do if you are unsure about the assessment.

The final part of the workshop was a practical exercise, where the TAs were asked to assess a fictitious student solution to a small introductory programming assignment in Python. All the participating TAs assessed the same solution, but they got different assessment instructions. Some TAs only got the lab instructions, some also got a list of requirements, and a third group, in addition, got a detailed grading rubric with possible errors and how to assess them. The TAs were first asked to grade the presented solution individually, according to their instruction. This was followed by a group discussion on their grading results and experiences with using the given assessment instructions. The exercise was after that concluded by highlighting that whether or not this particular student would have passed the assignment depended on by which criteria it was assessed and how the grader interpreted the requirements.

4 METHOD

All TAs, who were scheduled to be working during lab sessions in any of the introductory CS courses, were invited to participate in the workshop, altogether 94 TAs. This was a mix of newly recruited TAs and well-experienced TAs. It was voluntary for the TAs to participate, but they received their hourly salary for doing so. All course coordinators for CS1 and CS2 courses and the course coordinator of one more advanced CS course on algorithms, data structures, and complexity (an introductory course in

¹ The workshop material can be shared upon request and is also available as part of a TA training course (module 4): <https://canvas.kth.se/courses/19750/>

theoretical CS) provided us with contact lists to their TAs. In total, 50 TAs participated (39 during the fall and 11 during spring) and 44 did not.

4.1 Surveys

Pre-workshop surveys were distributed on paper to all workshop attendees and digital versions were distributed to non-attendees (control group). A post-workshop survey was distributed digitally in the final week of the semesters to all TAs who answered the pre-workshop survey. It was entirely voluntary for the TAs to answer the surveys. To allow for pre-post pairing, the answers were not anonymously collected but were anonymized after the pairing. All participants were asked to give their consent to be part of the study by declaring so in the last question of both surveys. The participants were informed of the purpose of the data collection and how the data would be handled.

The surveys consisted of 11 statements based on the findings in [29], (see S1-S11 in Fig. 1 and 2). The TAs were asked to rate how well they agreed with each statement on a 7-point Likert scale. If they lacked experience with what was in question, they were asked to skip the statement. The pre-workshop survey also included questions about prior TA training and work experience as a TA. In addition, workshop attendees were asked to rate if they benefited from the workshop (S12, see Fig. 3) in the post-workshop survey. Mann-Whitney U tests, non-parametric tests, were conducted to compare the distributions between the test group (workshop attendees) and the control group (non-attendees) on each statement for the pre-workshop survey and post-workshop survey answers, respectively [5, 16]. The hypothesis tested for each statement was that there was no difference between the groups. In addition, Wilcoxon matched-pairs signed-rank tests were carried out on the paired data (pre and post) for both groups separately [16], to test for differences between pre-and post-answers. Skipped statements were excluded.

4.2 Interviews

The interviews were conducted in parallel with the survey data collection, allowing for a triangulation between the data [4]. To prevent any interference, the survey data were analyzed after all interviews were completed. The interviews were semi-structured with open questions. This was chosen to allow the TAs to reflect and talk about their personal experiences and allow for follow-up questions. The TAs were asked to reflect and describe their practice and TA role. For example, which work tasks they have, how they carry out the assessment, and how they decide the grade of a students' solution. See [30] for the complete interview guide and a more detailed analysis of the interviews.

In total, seven TAs were interviewed, four workshop attendees, and three non-attendees. Six of the interviews were conducted face-to-face and one through an online video chat. The interviews lasted between 30 minutes to an hour. It was completely voluntary for the TAs to participate. All interviewees agreed to have their interviews audio-recorded and gave their consent to be part of this study. The interviews were all carried out by the first author of the paper, who had conducted part of the workshops but was not a course coordinator for any of the studied courses. As a small token of appreciation, all interviewees received a cinema ticket gift card.

The interview data were analyzed using the survey data statements as a framework to seek explanations and further detail. That is, the transcripts were coded corresponding to the statements and summarized into themes and/or explanations of them.

5 RESULTS

In this section, we first present the results from the survey (descriptive) and statistical test that were carried out. In the second subsection, we present the results from the interview data, which offer further explanations of the survey results.

5.1 Survey Results

All 50 workshop attendees and 18 non-attendees filled out the pre-workshop survey. The post-workshop survey was answered by 31 of the workshop attendees and 14 of the non-attendees. The Mann-Whitney U tests on the pre-workshop survey answers for the 11 statements, on a 95% confidence level, showed a significant difference between the groups, on one statement (S11: *If there are no assessment criteria, the lab instructions tend to be detailed enough for me to assess whether a presentation should be*

accepted or not.), where the test group agreed with the statement to a larger extent compared to the control group (p-value = 0.028). An equal proportion of the TAs from both groups had attended prior TA training. However, the TAs in the control group were slightly more experienced (see Table 2).

Table 2. Prior teaching experience and training for the workshop attendees and non-workshop attendees.

Previous TA training experience	Workshop attendees percentage (number) n = 50	Non-workshop attendees Percentage (number) n = 18
None - never attended any training.	22% (11)	22% (4)
Had attended at least one TA training session before	78% (39)	78% (14)
Previous experience being a TA	Workshop attendees	Non- workshop attendees
None - first semester	42% (21)	28 % (5)
One course during 1 semester.	12 % (6)	28% (5)
One/multiple courses, 2 semesters.	20 % (10)	11% (2)
One/multiple courses, 3-4 semesters.	10% (5)	0% (0)
One/multiple courses, 5-6 semesters.	8% (4)	22 % (4)
One/multiple courses, >7 semesters.	8% (4)	11 (2)

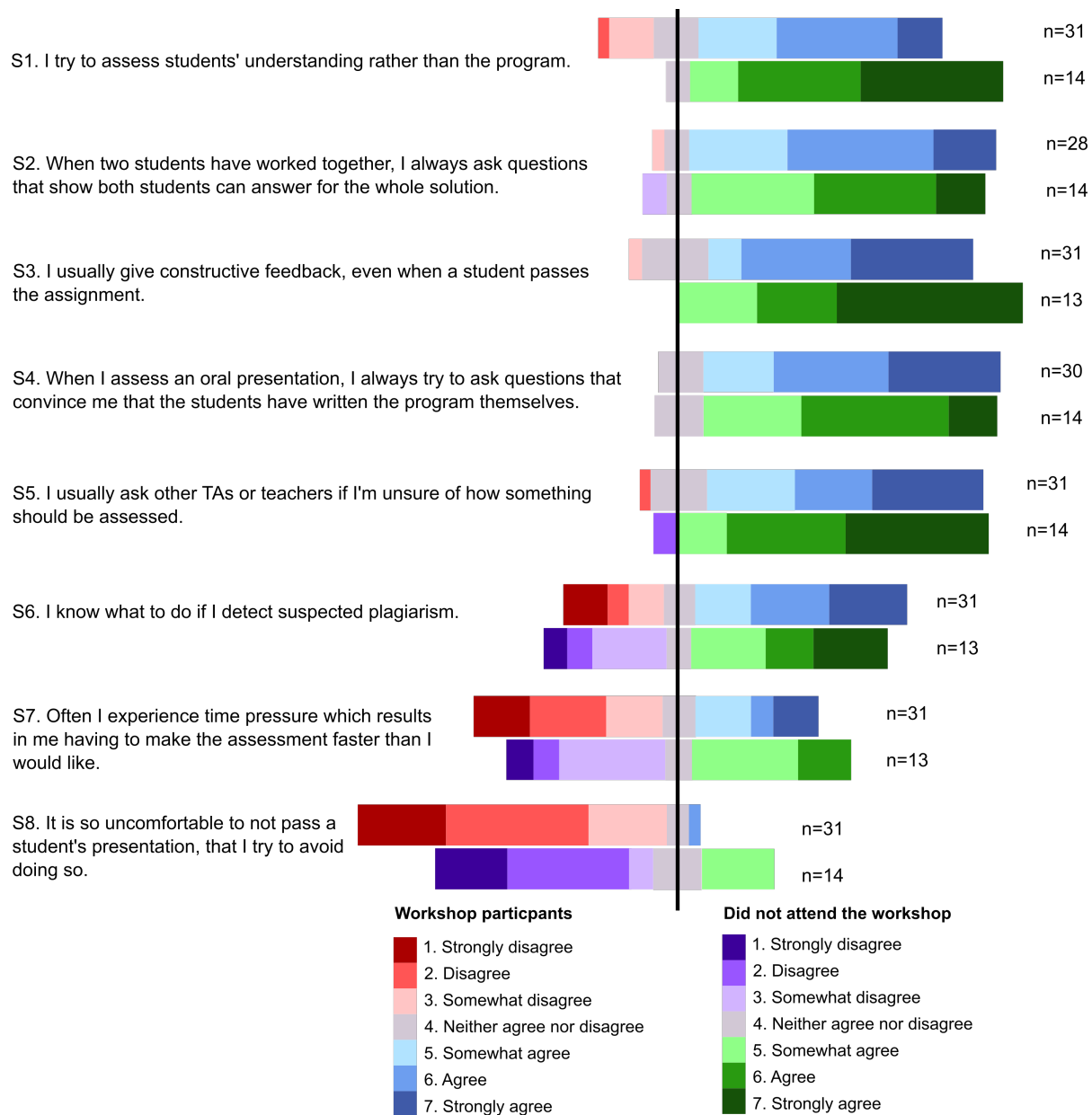


Fig. 1: TAs' answers to the asked statements S1-S8 from both groups in the post-workshop survey, in percentage for each group.

Having prior experience was also given as an explanation for not attending the workshop, but the main reason given was scheduling conflicts. Eight statements (see Fig. 1, S1-S8) regarding how the TAs conduct the assessment related to RQ1 (how TAs conduct and handle the assessments of students' lab assignments). This includes aspects that had previously been identified as challenging [29], such as having to fail a student, time pressure, and how to handle suspected plagiarism.

Three statements were asked to address RQ2 (what material and instructions should be provided), since it has previously been found that TAs had experienced a lack of instructions [29] (see Fig. 2, S9-11).

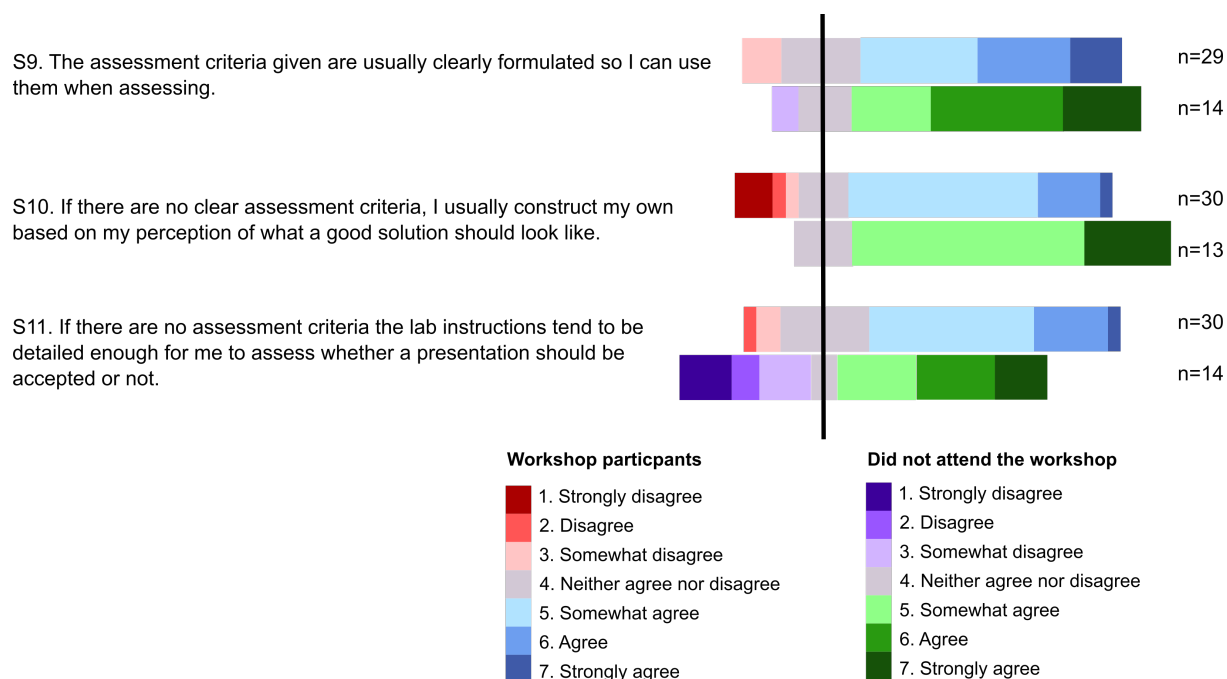


Fig. 2: TAs' answers to the asked statements S 9-S11 from both groups in the post-workshop survey, in percentage for each group.

The Mann-Whitney U test, testing the hypothesis that there were no differences between the test group and the control group, showed that there was a significant difference at the 95 % confidence level, only on the first statement (S1) in the post-workshop survey, where the test group to a greater extent disagreed with the statement compared to the control group (p-value 0.014).

To address RQ3 (if a training workshop can change the TAs' practices and experiences) all post-workshop survey answers on the 11 statements were paired with the corresponding pre-workshop survey answers, from both groups. A Wilcoxon matched-pairs signed-rank test, testing the hypothesis that there was no difference between the pairs on a 95% confidence level, showed a significant difference on three questions for the test group, but no significant difference on any of the questions for the control group (Table 3 and Fig. 3). In addition, the TAs attending the workshop were asked to rate how much they benefited from the workshop, see Fig. 4.

Table 3: Wilcoxon matched-pairs signed-rank tests results.

Statement	p-value
S4. When I assess an oral presentation, I always try to ask questions that convince me that the students have written the program themselves.	0.020
S5. I usually ask other TAs or teachers if I'm unsure of how something should be assessed.	0.021
S11. If there are no assessment criteria the lab instructions tend to be detailed enough for me to assess whether a presentation should be accepted or not.	0.001

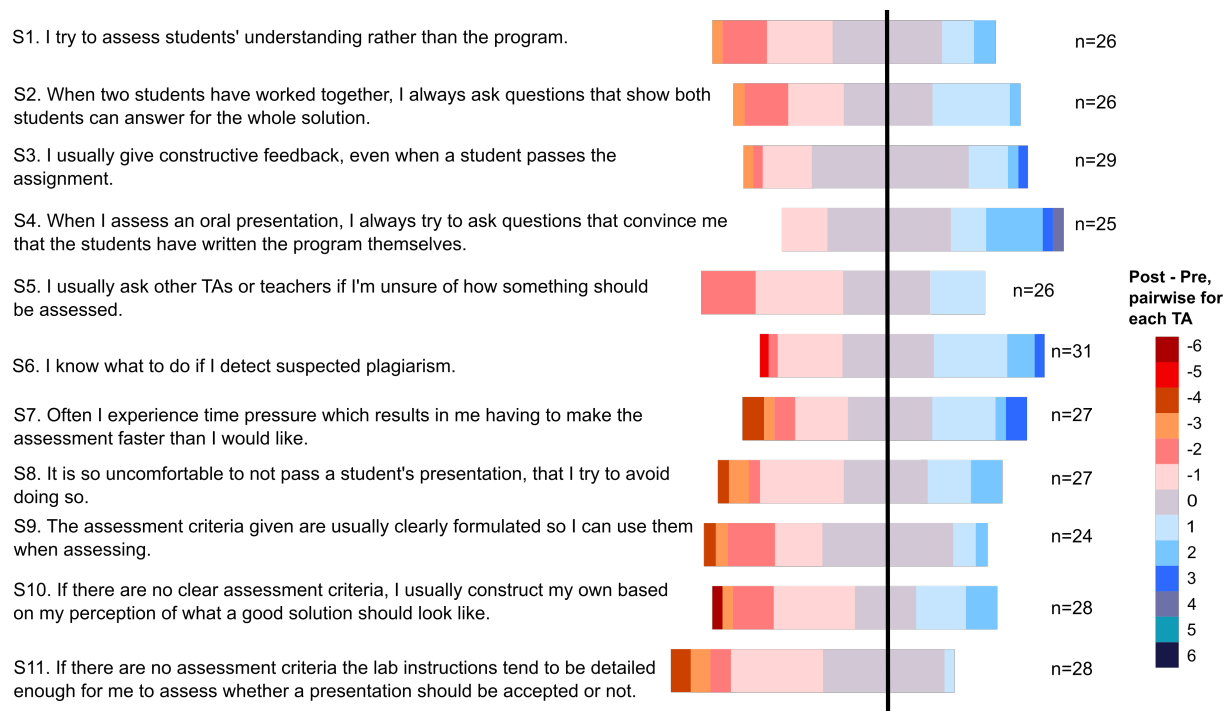


Fig. 3: Post-workshop survey ratings subtracted by pre-workshop survey ratings for the test group. E.g. a positive difference means that the respondent stronger agreed with the statement in the post-test compared to the pre-test.

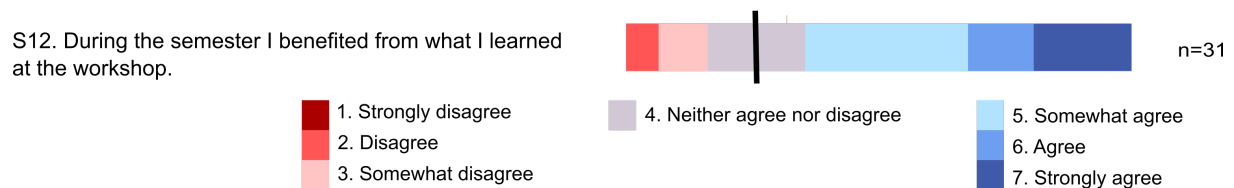


Fig. 4: TAs' answers to statement about the workshop.

11.1 Interview Results

In this section, a summary of the interview findings is presented, and key aspects are also highlighted by exemplary quotes that have been translated to English by the first author.

The interviewed TAs all stated that an important part of the assessment was that the students could explain how their code worked, that the students understood what they had written. This was explained as trying to assess the students' understanding, rather than solely the functionality of the program (as asked about in S1). Some TAs, however, said that they relied solely on the lab instructions, which often only included the functional part of the code. If the students could not answer questions regarding concepts, the TAs would instead explain the answers to the students. If the students could not answer simple questions regarding their code, the TAs might suspect the students had copied it, and the questioning of the students then had to do with checking for plagiarism (as asked about in S4). Below follows an example quote:

"If you are doubtful whether or not they have written it [the program] themselves, it is easier to say that it is not approved. But then I personally try to strive after ... if at one point they do not really understand anything, to explain that to them and then ask some follow-up questions to see that they have understood. So the point, somehow, is that you learn too." - Workshop attendee, 1.5 years of experience as a TA

The interviewed TAs all explained that they would contact the course coordinator if they suspected plagiarism, even if some of the TAs were a bit uncertain if that was the correct strategy (in S6, we asked if the TAs knew what to do if they detected suspicious plagiarism). In line with the survey answers, the interviewed TAs said that if the students presented their solution in a pair, they try to ask equally hard questions to both students (S2). The interviewed TAs also said that they usually give the students

feedback even when the students' solutions are accepted, given that they have the time to do so (if they offered students with feedback was asked about in S3 and if they experienced time pressure in S7).

"I think feedback on code is something that is really missing and that is why I try to give it all the time."
– Non-workshop attendee, 2 years of experience as a TA

Regarding the large spread in answers on the statement regarding time pressure (S7), the interview answers revealed that the time pressure could differ much between courses and sessions. Experienced TA also stated that they learned to handle the stress.

"Sometimes the lab sessions can be quite stressful because the queues are long. I experienced that stress much during the first year as a TA, after that I let it go." - Workshop attendee, 3 years of experience as a TA

The interviewed TAs stated that they could ask for advice from their TA colleagues or from their course coordinator if they needed it, as suggested in the survey answer (S5). They stated that they had experienced a greater need for help when they were newly employed or started to work in a new course. The more experienced TAs said that they were the ones other TAs turn to for help.

"Now there are other assistants who come and ask me and so on. So that is a solution to deal with that [being uncertain with an assessment], to go to another assistant." - Workshop attendee, 3.5 years of experience as a TA

Some of the interviewed TAs had been a bit uncomfortable failing a student (a factor asked about in S8) but would try to allow the students to make corrections and present again.

"It is very rare that it is like 'you have not managed anything, you are screwed'. But rather 'you don't really understand these things, so you just need to read a little about it and try to understand it, once you done that join the queue again and write my name and I'll come back.'" - Workshop attendee, 1.5 years of experience as a TA

The interviewed TAs explained that it is hard to fail a student's solution for something that is not clearly written in the lab instructions. How detailed the instructions were could differ between courses and assignments, which align with the spread of results in the survey (see S9, S11). If the instructions were unclear, the TAs could construct criteria by themselves (as also stated in S10), typically independently. The interviewed TAs stated to typically not get the chance to discuss how the given instructions/criteria should be interpreted and that not all course coordinators had the time or willingness to construct and use grading criteria.

"It actually depends. Some courses have very good templates for how the grades should be given. One course has an extremely good template, not only for the project work but also for the various lab tasks it says 'this should be included, this you can review a little less.' [...] In the other course, there were very bad instructions. They had not even passed a spell check or anything, so it was very much that I had to go with what I thought [was right]." - Non-workshop attendee, 2 years of experience as a TA

Regarding RQ3, the interview findings suggested that new TAs found the workshop more helpful than experienced TAs. Both workshop attendees and non-attendees stated that they would also have liked to be offered introductory training.

"I think, above all, that it [the workshop] had been valuable when you start out. Because there was no such thing when I started and then you had to form your own opinions." - Workshop attendee, 1.5 years of experience as a TA.

12 DISCUSSION

The section is divided between the three RQs, followed by a section on limitations and threats to validity, and concluded by a discussion of implications and future plans.

12.1 RQ1- How Do TAs Conduct the Assessments?

The trends in the results (see Fig. 1, S2-S5) show that our TAs ask questions to both students within a pair, give constructive feedback, ask questions that convince them that the students have written the programs, and ask for help if they are unsure. All of which indicates that the TAs had similar practices,

well in line with what is expected by the course coordinators. Most TAs also stated that they were comfortable failing a student (S8), a necessary for grading, but somewhat contradictory to the previous finding in [29]. The answers to S7 showed that time pressure had been experienced, but to which extent varies, which is in line with prior research [29]. The most alarming finding in the TAs' practices is that many TAs did not know how to handle suspected plagiarism (S6), which was included in the workshop but yielded no effect. Handling suspicion of cheating has been identified as one of the challenges TAs experience with conducting assessments [29, 32], and this is an aspect that should be further addressed in future iterations.

The difference between the two groups, that workshop non-workshop attendees stronger agreed with S1 (that the assess the students understanding rather than the program's functionality), might seem like a step in the wrong direction. However, in many of our lab assignments, only requirements for the functionality of the code is given. Thus, this result is actually in coherence with the instructions. However, it indicates that it could be ambiguous for the TAs what the focus of the assessment is. Typically, the assignments can be categorized as a high cognitive processes level [1], namely *create*, and on a conceptual or procedural level in the knowledge dimension. Having the students explain solely the functionality, rather falls under the cognitive processes level *understand*, or *remember*, while the emphasis on the problem solving and development of the code, falls under the level *create*. These results highlight the importance of communication and defining what should be assessed and how, which is particularly important when the assessment is conducted by multiple TAs and not a trained instructor.

12.2 RQ2 - What Materials and Instructions Are Needed?

We have evidence to support that only providing lab instructions as the basis for the TAs to carry out the assessment is insufficient from the TA perspective, similar to previous findings [29] (S11). Grading criteria or grading rubrics are currently not always used (as stated in S10-S11), and in the interviews, our TAs described it to be due to the unwillingness, as in [27]. However, when grading criteria are provided, the criteria are experienced as clear by the TAs (see responses to S9). The TAs from both groups also stated that they would construct their own grading criteria if given insufficient instructions. From a reliability point of view, it is an issue if the TAs construct their own criteria, especially since the interview results show our TAs work independently, as was also found in [29, 30]. It has previously been shown that the reliability of the assessments was higher when the TAs carried out the assessment together [15] and that social aspects, such as collaborating with other TAs, affect job satisfaction [24]. Our TAs did state that they asked for help if they were unsure about an assessment (S5), but to also facilitate discussion before the assessments, could address any ambiguous phrasing in the instructions or criteria given, as also suggested regarding the use of grading rubrics in [27]. The TAs play a vital role in the assessments and to achieve constructive alignment [2] and the material needs to be understandable to both students and TAs.

12.3 RQ3 - Does a Training Workshop Make a Difference?

We found three significant differences between post-workshop and pre-workshop answers in the test group on S4, S5, S11. No differences were found in the control group, which implies that these changes are not caused by changes in the material/instructions or by only gaining TA experience. Similarly to the changes that were found for faculty members who participated/did not participate in pedagogical training [26]. We were, however, surprised by the fact that the TAs perceived the grading instructions as less sufficient after the workshop (S11). The workshop attendees could, however, have become better at identifying what was missing, after being exposed to a wider variety of examples during the training. Hopefully, the TAs were also empowered by the workshop training to request more clarifications or additional material, if needed. The S5 results (showing that TAs were less inclined to ask for help) imply that the workshop attendees did not have to seek help as often. As brought up during the interviews, the more experienced TAs were instead more often asked for advice by others, similar to mentorships as have also been suggested as a training method [25]. Since the workshop addressed the previously identified issue on how to distinguish that the student wrote the code from that the students can explain how the code works [29], that is thought to be the reason for the positive change in S4. We are, however, disappointed with the fact that we saw no improvement regarding if the TAs knew how to handle suspected plagiarism (S6), which was specifically addressed during the workshop.

12.4 Limitations and Threats to Validity

Since participation in the training was voluntary, one might suspect that the TAs belonging to the test group were more eager to develop their TA practices, compared to the control group. However, as shown in Table 2, an equal proportion of the TAs from each group had attended prior TA training. The non-attendees had slightly more work experience, but the pre-workshop answers were similar, which would make the post-workshop answers comparable. Even though it would have been desirable with a random selection, ethically, we found it important to give all TAs the same opportunity to participate in the TA training workshop.

The statements that were used in the surveys were constructed based on the findings in [29] but have not been validated further. We cannot completely rule out the possibility that some TAs answered the survey, thinking there was a “right” answer. However, what they answered, or if they answered, had no implications for the TAs, as the surveys were neither graded nor shared with their respective course coordinators. The triangulation of the survey and the interview data also strengthens the trustworthiness of the results [4]. The sample size and size of the intervention (conducted at one institution) is also a limitation to the study, and it would be interesting to run the workshop at another institution to evaluate the training further and explore potential differences in TAs' practices.

12.5 Implications and Future Plans

The results show that the instructions and material given to our TAs could be improved and that the TAs could benefit from working more together. We would, therefore, advise course coordinators to provide their TAs with grading criteria for each lab assignment and to facilitate communication between the TAs. Based on the results from this study, we have made some changes to the workshop, allowing for a longer and more detailed discussion regarding plagiarism and stress, and continued to offer the workshop. During 2020 the workshop became part of a mandatory TA training course, which also includes training in other aspects such as how to give a tutorial and tutor students. We have already collected some evaluation data from the course participants and will present the results in a forthcoming study [31].

CONCLUSIONS

We conclude that there is a need for TA training. The TAs who attended our workshop were less inclined to ask for help, experienced that the lab instructions lacked necessary details, and were more inclined to ask questions that convinced them that the students have written the program themselves. For the TAs to be successful and able to carry out valid and fair assessments, it is also important that the instructors provide their TAs with sufficient material. Based on these results, we recommend course coordinators to develop grading criteria and grading rubrics for the lab assignments and educate the TAs on how to use them.

REFERENCES

- [1] Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R.,Raths, J., & Wittrock, M.C. (2001), “A taxonomy for learning, teaching, and assessing: A revision of Bloom’s Taxonomy of Educational Objectives” New York: Longman.
- [2] Biggs, J., (1996), “Enhancing teaching through constructive alignment”, *Higher Education*, 32: 347–364.
- [3] Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, WH, & Krathwohl, D.R., (1956), “Taxonomy of educational objectives: The classification of educational goals”, *Handbook 1: Cognitive domain*. New York: David McKay.
- [4] Borrego, M., Douglas, E.P., Amelink, C.T, (2009), “Quantitative, qualitative and mixed research methods in engineering education.” *Journal of Engineering Education*. 98(1), pp. 53–66
- [5] Brace, N., Kemp, R., Snelgar R. (2016), “SPSS for psychologists - And everybody else”. Red Globe Press.
- [6] Brent R., Maners, J., Raubenheimer, D., and Craig, A., (2007), "Preparing undergraduates to teach computer applications to engineering freshmen" In 2007 *Frontiers in Education Conference (FIE)*.
- [7] Danielsiek, H., Vahrenhold, J., Hubwieser, P., Krugel, J., Magenheimer, J., Ohrndorf, L., Ossenschmidt, D. and Schaper, N. (2017), “Undergraduate teaching assistants in computer science: Teaching-related beliefs, tasks, and competences”, In 2017 *IEEE Global Engineering Education Conference (EDUCON)* pp. 718–725.

- [8] Dechenne, S. E., Koziol, N., Needham, M. and Enochs, L., (2015). “Modeling Sources of Teaching Self-Efficacy for Science, Technology, Engineering, and Mathematics Graduate Teaching Assistants,” *CBE-Life Sciences Education*, vol. 14, no. 3.
- [9] Decker, A., Ventura, P., and Egert, C., (2006), “Through the Looking Glass: Reflections on Using Undergraduate Teaching Assistants in CS1”. In *Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education (SIGCSE' 06)*. ACM, New York, NY, USA, pp. 46–50.
- [10] Dickson, P. E., Dragon, T., and Lee, A. (2017), “Using Undergraduate Teaching Assistants in Small Classes”, In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, pp. 165–170.
- [11] Dixson, D.D. and Worrell, F.C., (2016), “Formative and Summative Assessment in the Classroom.” *Theory Into practice*, 55(2), pp.153–159.
- [12] Enström, E., Kreitz, G., Niemelä, F., Söderman P., and Kann, V., (2011), "Five years with kattis — Using an automated assessment system in teaching", 2011 *Frontiers in Education Conference (FIE)*, 2011, pp. T3J-1-T3J-6, doi: 10.1109/FIE.2011.6142931.
- [13] Estrada, F. J., and Tafliovich A., (2017), “Bridging the gap between desired and actual qualifications of teaching assistants: An experience report”, *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*.
- [14] Forbes, J., Malan, D., Pon-Barry, H., Reges, S. and Sahami, M., (2017), “Scaling introductory courses using undergraduate teaching assistants,” *Proceedings of the 2017 ACM SIGCSE Technical Symposium on computer science education*, pp.. 657–658, 2017.
- [15] Harrington, B., Ahmadzadeh, M., Cheng, N., Wang E. H., and Efimov, V., (2018), “TA marking parties: Worth the price of pizza?”, In *Proceedings of the 2018 ACM Conference on International Computing Education Research*, pp. 232–240.
- [16] Howell, D. C., (2012), “Statistical methods for psychology” Wadsworth Publishing Company; International Edition, 8th edition.
- [17] Kay, D.G., (1995). “Training computer science teaching assistants: a seminar for new TAs”, *Proceedings of the 1995 ACM SIGCSE Technical Symposium on computer science education*.
- [18] Leyzberg, D. and Lumbroso, J., (2017), “Nailing the TA interview: using a rubric to hire teaching assistants”, *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*.
- [19] Lorås. M., (2020), “From teaching assistants to learning assistants – lessons learned from learning assistant training at Excited”, *Læring om læring* 4 (1).
- [20] Maruti, F., Rodgers, K., Jung, H., Moon, A., and Diefes-Dux, H., (2013), “Factors that help and hinder teaching assistants’ ability to execute their responsibilities,” *Proceedings of the 120th ASEE Annual Conference and Exposition*.
- [21] Mirza, D., Conrad, P. T., Lloyd, C., Matni, Z. and Gatin, A., (2019), “Undergraduate Teaching Assistants in Computer Science: A Systematic Literature Review”, In *Proceedings of the 2019 ACM Conference on International Computing Education Research - ICER '19*, pp. 31–40.
- [22] Moon, A.C., Jung, H., Marbouti, F., Rodgers, K.J., and Diefes-Dux, H., (2013), “Undergraduate and graduate teaching assistants' perceptions of their responsibilities - Factors that help or hinder”. In *Proceedings of 2013 IEEE Frontiers in Education Conference (FIE)*, pp. 1576-1578.
- [23] Muzaka, V. (2009), “The niche of graduate teaching assistants (GTAs): Perceptions and reflections,” *Teaching in Higher Education*, vol. 14, no. 1, pp. 1–12.
- [24] Patitsas, E., (2012), “A case study of environmental factors influencing teaching assistant job satisfaction”, In *Proceedings of the ninth annual international conference on International computing education research (ICER '12)*. Association for Computing Machinery, New York, NY, USA, 11–16. DOI:<https://doi.org/10.1145/2361276.2361280>
- [25] Patitsas, E., (2013), “A case study of the development of CS teaching assistants and their experiences with team teaching”, In *Proceedings of the 13th Koli Calling International Conference on Computing Education Research*, pp. 115–124.
- [26] Postareff, L., Lindblom-Ylänne, S. & Nevgi, A. (2008), “A follow-up study of the effect of pedagogical training on teaching in higher education”. *High Educ* 56, 29–43. <https://doi.org/10.1007/s10734-007-9087-z>
- [27] Reddy, Y. M., and Andrade, H. (2010), “A Review of Rubric Use in Higher Education,” *Assessment & Evaluation in Higher Education*, vol. 35, no 4, pp. 435–448.
- [28] Ren, Y., Krishnamurthi, S., and Fisler, K., (2019), "What Help Do Students Seek in TA Office Hours?" In *Proceedings of the 2019 ACM Conference on International Computing Education Research (ICER' 19)*.

- Association for Computing Machinery, New York, NY, USA, 41–49. <https://doi.org/10.1145/3291279.3339418>
- [29] Riese, E., (2018), "Teaching Assistants' Experiences of Lab Sessions in Introductory Computer Science Courses", 2018 IEEE Frontiers in Education Conference (FIE), 2018, pp. 1-5, DOI: 10.1109/FIE.2018.8659243.
- [30] Riese, E., and Kann, V. (2020), "Teaching Assistants' Experiences of Tutoring and Assessing in Computer Science Education," 2020 IEEE Frontiers in Education Conference (FIE), 2020, pp. 1-9, DOI: 10.1109/FIE44824.2020.9274245.
- [31] Riese, E., and Kann, V., (2022, forthcoming), "Training Teaching Assistants by Offering an Introductory Course", In Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - SIGCSE '22, DOI: <https://doi.org/10.1145/3478431.3499270>
- [32] Riese, E., Lorås, M., Ukrop, M., and Effenberger, T., (2021). "Challenges Faced by Teaching Assistants in Computer Science Education Across Europe". In Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE' 21). Association for Computing Machinery, New York, NY, USA, 547–553. DOI:<https://doi.org/10.1145/3430665.3456304>
- [33] Roberts, E., Lilly, J., and Rollins, B., (1995), "Using undergraduates as teaching assistants in introductory programming courses: an update on the Stanford experience." SIGCSE Bull. 27, 1 (March 1995), pp. 48–52.
- [34] Sadler, D. R., (2009), "Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning", In Assessment, Learning and Judgement in Higher Education, Dordrecht: Springer Netherlands, pp. 45-64.
- [35] Ukrop, M., Švábenský, V., and Nehyba, J., (2019), "Reflective diary for professional development of novice teachers," In Proceedings of the 50th ACM Technical Symposium on Computer Science Education - SIGCSE '19, pp. 1088–1094.
- [36] Vihavainen, A., Vikberg, T., Luukkainen, M., and Kurhila, J., (2013). "Massive increase in eager TAs: experiences from extreme apprenticeship-based CS1" In Proceedings of the 18th ACM conference on Innovation and technology in computer science education (ITiCSE' 13). Association for Computing Machinery, New York, NY, USA, The feedback was solely positive, confirming the need for the workshop, its content, and its structure. pp.123–128.