Equity-Guided Tutoring: a Scalable, TA-Run Approach

Daniel Patterson
dbp@dbpmail.net
Northeastern University
Boston, MA, USA

Josh Torre
torre.jo@northeastern.edu
Northeastern University
Boston, MA, USA

Zoey Guo
guo.zoe@northeastern.edu
Northeastern University
Boston, MA, USA

Thomas McBride
mcbride.t@northeastern.edu
Northeastern University
Boston, MA, USA

Abstract

We report on a prototype small tutoring program run entirely within the existing undergraduate teaching assistant (TA) staff of a large (around 600 students) first semester Computer Science class with an equally large (around 80) staff of undergraduate TAs. The tutoring team, led by the second, third, and fourth authors (all undergraduate TAs), identified students who could benefit from the service by analyzing existing course assessments (homework, labs, and exams), some of which were slightly modified to aid this task.

The service was intentionally not advertised nor available by student request to avoid being dominated by proactive students who benefited from the traditional office hours offered by the course. The tutoring leads matched a team of around 15 TAs with students identified to be at risk of failing the class. 92 one-on-one tutoring sessions were conducted with the matched students over the course of the semester. Of these students, the majority passed the class, and their exam scores improved significantly more than the class population as a whole. Additionally, qualitative feedback demonstrates the program's positive impact on the students.

CCS Concepts

• **Social and professional topics → Computer science education**.

Keywords

equity, tutoring, teaching assistants

1 Background

In large introductory classes, a certain percentage of students do not pass every semester [2]. Historically, pass rates at our institution have been at or above the worldwide average. However, with course failure-to-retain rates rising into the double digits, an obvious question arises: are there interventions that could improve these numbers? While there will always be a certain number of students who do not pass any class, the hypothesis that drove this prototype program is that some of those students fell *slightly* behind at some point and were unable to catch up due to the cumulative nature of the class.

Our class, like most introductory computer science classes, relies and builds upon skills developed earlier in the semester as the semester unfolds. While this feels necessary for beginning programming courses, it does mean that difficulty early in the semester can cause increasingly significant issues for students. This "left behind" hypothesis posited that limited amounts of extra help, specifically targeted at students that may not even know they are in the position we identify them to be, could make a significant difference. While the overall number of students reached in this prototype program was relatively small—only a few dozen—the time commitment was less than one hundred hours of one-on-one tutoring time, out of over ten thousand hours spent collectively by the undergraduate TA team over the semester.

2 Related Work

While undergraduate TA programs are extensive and many describe what they offer as tutoring, in the literature [4], most of these "tutoring" programs operate on a first-come-first-serve basis at fixed hours. We consider this as "office hours", and contrast it to our approach in §4.1. While these are important, we believe it orthogonal to the program addressed in this report.

Akhmetov et al. [1] describes a "Lost student support" program implemented as part of a larger organizational structure in their large introductory courses, which proactively identifies students similar to our program. Their assistance for identified "lost students" is limited to a single 15-30 minute meeting, primarily targeting students who have a small logistical challenge, rather than our program, which aims to rectify content issues for students. With our program offering three to four one-hour weekly sessions, we also provide significantly more time with the student.

There are also mentoring programs which are not tied to any particular courses. While these certainly have a positive impact [3] because they exist outside of courses, they are likely to intervene reactively in response to poor grades in courses, thus requiring significantly more time and work for both the student and mentor to get the student back on track. In contrast, our program requires very

little additional overhead through early intervention and redirection of existing course resources.

3 Intervention

At the beginning of the semester, a team of undergraduate teaching assistants (TAs) led by the second, third, and fourth authors were tasked with: A. identifying students who were at risk of failure, likely inadequately served by existing course resources (lectures, labs, office hours) and B. offering one-on-one tutoring sessions to those individuals.

One key principle was that this program **would not be advertised**. It did not show up on our course website (unlike our extensive office hour schedule), and whether a student *asked* for tutoring was not a factor in determining whether tutoring was assigned.

If we did decide based on student interest, we would end up helping students who were proactively trying to do as well as they could in the class, but not necessarily those that *needed* the help most. To maximize our impact, we wanted to ensure that we, the instructional staff, were deciding who to help based on objective measures of their performance relative to other students in the course.

3.1 Summary of Program

Once identified, students were offered one-on-one tutoring sessions with one of the TAs on the tutoring team. The second key principle was that **help would be limited**. The sessions were one hour long, offered weekly and generally limited to a total of three or four sessions. This duration allowed enough continuity to help with concrete material and establish strategies on how to approach the coursework. We also wanted there to be a way for tutors to see if those strategies were working and adjust as needed, which would be difficult to support with only one or two sessions.

The tutoring sessions were always intended to end, which allowed TAs to move on to different students, ensuring that the students were eventually succeeding based on their own engagement with the course material and course resources.

The third key principle was that **tutees would be chosen to maximize impact**. This led to two phases of the program, with two different identification criteria. At the beginning of the semester, we chose students who were struggling the most in class based on concrete assessment. We designed a lab a few weeks into the semester that was review material carried out solo rather than in pairs, which most students completed quickly. Those that struggled were directly referred to the tutoring team.

Towards the latter half of the semester, the focus changed to students who had performed well at the beginning of the semester but whose performance had dropped significantly towards the middle. These students were at risk of failure if their performance continued on the same trajectory, but they were not doing the worst in class. Students performing poorly on nearly all the course content midway through the semester were unlikely to change their trajectory even if they received three to four hours of one-on-one help.

Students were "off-boarded" from the tutoring program based on their success on one of the three midterms spread evenly throughout the semester, their own confidence with the material, or after all four tutoring sessions were completed. While some were still struggling at the point of off-boarding, the time constraints of the TAs involved in the program combined with our operating hypothesis that a limited amount of help could still make a difference led us to proceed with off-boarding.

3.2 Identifying Students

We used four primary mechanisms to identify students.

Course staff could flag students who they believed would benefit from tutoring. At the beginning of the course, before we had many course assessments to use, this helped identify students that didn't seem to be keeping up with their peers. Instructors and TAs determined this based on their performance in labs, where TAs directly observed the students.

We also designed two labs, which we called "calibration labs", run in the third and seventh weeks of class. These were intentionally a review of the previous weeks of material, and were meant to be solvable in less than half the allocated lab time for most students. Unlike all other labs, these were completed individually, and the few students who were still working on them at the end of the lab period were directly referred for tutoring. The labs also served as review for the first and second midterms, which occurred shortly after.

The first midterm took place about a month into the class, and anyone with a failing grade on the exam was referred to the tutoring team. This first midterm was also our first real opportunity to see that this program was working, as many of the students identified early in the semester passed the exam, sometimes with very good grades!

The second midterm, which occurred around two months into the class, was used to the measure success of the program and off-board students, but also to identify new potential tutees. Those identified were students who performed significantly worse on the second midterm than the first, indicating that they had fallen behind relatively recently.

3.3 Working With Students

Tutoring sessions were typically scheduled on a weekly basis, with each student allowed up to four one-hour sessions per semester. This limit could be extended to five sessions if deemed necessary. Each session focused on addressing the student's challenges with concepts from previous lectures. Activities included reviewing lecture notes, solving practice problems, and creating examples to demonstrate the application of specific concepts and tools taught in class. Rather than providing a similar function to office hours (where students mainly ask for help on assignments), tutoring sessions solely focused on solidifying the students' general understanding of the course material.

For sessions held shortly before an exam, the focus shifted to working through practice exams and discussing strategies for tackling potential problems. After each session, both the TA and the tutee were required to complete feedback forms related to the tutoring program. Tutees provided feedback on the session's effectiveness, shared any complaints or suggestions, and indicated their interest in future sessions. TAs evaluated the tutee's progress and documented the topics covered during the session.

3.4 Off-boarding

Another principle we decided at the beginning was that tutoring would be a limited service. This was partly due to our own capacity, as there was a limit to how many hours that could be dedicated to this program on the scale of a 600 student class with only 15 TAs who had other duties and responsibilities. Additionally, this intervention was only intended to bring a student back to where they could use regular course resources: e.g., daily traditional office hours with usually more than 1 TA available at a time.

After a few weeks of running the program, we decided that our limit would be 4 ± 1 sessions, which seemed like a good balance between giving students a solid amount of help and allowing the TAs to move on to new students that were identified.

In addition to off-boarding based on the number of sessions, we also off-boarded students if they scored 90% or higher on a midterm or if they decided they did not need the service anymore.

3.5 Benefits of Program

Since TAs are a part of the course staff, they understand the material, pace of course, and assessment very well. This makes them well suited to tutor the course content one-on-one with students. In addition to understanding the course material, TAs also have the benefit of already existing on course payrolls. Our prototype involved less than a quarter of the TAs, and each only spent an hour or two per week on this program while still making a noticeable impact. This meant that the program had low logistical overhead to run. Another advantage of the program being run almost entirely by TAs is that it can easily scale up or down. The only management that the first author (lead instructor for the course) did, after assigning the second, third, and fourth authors to run the team, was to consult on some of the details about assigning work, off-boarding students, identifying students, and constructing the "calibration labs".

3.6 Challenges

Our main challenge was ensuring that the TAs who volunteered to tutor truly followed through in working with their tutees. Given this was a new program, none of them were familiar with it, and we did not have existing training materials for them. While the skills they used in traditional office hours mostly translated, some of the logistics of communication individually and scheduling sessions, both of which they did not have to do for office hours, proved challenging.

Initially, the team leads assigned tutors a student to schedule tutoring sessions with, expecting the tutor to reach out to the student and set up sessions. This turned out to be less than ideal, as it was often hard to tell exactly who was meeting with students, and whether they failed to meet because the student never responded or the TA did not follow up.

Of the 15 TAs who volunteered to be on the team, two never held a single tutoring session, and even though all others help some, the majority of sessions were held by a few TAs. Eventually, we switched the system to have one of the team leads (themselves TAs) manage each TA, and handle primary communication to set up the sessions. This made it much easier to determine when sessions occurred and ensured tutor accountability. This change was positively received by the tutors who, when asked what could be done to make the program better, requested continuing this style of delegation.

4 Alternatives

Currently, there are two main alternatives for students: office hours and external "peer" tutoring, for our students via a program called Knack. Students can find paid tutoring services, but it is likely they suffer the same issues as external peer tutoring, so we don't address them separately. Each alternative has significant drawbacks, which are detailed in the following sub-sections, where we address how our program differs.

4.1 Office Hours

Office hours are the main resource offered by the course, and are incredibly important, but they are intentionally designed to be equally available to all students, regardless of how well they are doing in the class. While we can expect students at the very top of the course to not use office hours much (or at all), there is a spectrum of students who use them, and there is no easy way for TAs holding office hours to allocate help to students based on need.

There are two primary reasons for this. First, office hours are often carried out in large shared rooms, with public queues identifying how many students are waiting. While at certain times of day or days of the week they may be quiet enough to allow significant one-on-one help, at others, TAs must move through students steadily to not draw the ire of others in the queue.

Secondly, until a TA is interacting with a student, there is no way for them to know how much help they need, and therefore allocating help based on need is an impossible task: perhaps they spend half an hour helping one student, who seems like they need a lot of help, only to have the next student need even more.

While it is certainly possible to attempt to establish a form of triage for office hours, and sometimes this is done in order to find common questions, it likely would not work to allocate help based on need. By attending office hours, students show that they are struggling in the course and deserve to be helped!

The point of the program described in this paper is not that we disagree with the idea of helping all students. Instead, we see our program existing to help students progress from struggling with course resources to using them effectively. Clearly, such a program is not necessary for someone for whom office hours are already working.

Our internal tutoring program resolves these issues by identifying and only working with students who have demonstrated the most need. One-on-one time, outside of office hours, can be set up without worrying about not allocating time most efficiently or being "unfair" during office hours.

4.2 External-to-course peer tutoring

The other resource that is available to students are third-party "peer tutoring" systems. Students can sign up to tutor specific subjects one-on-one, and other students enrolled in those courses can then sign up to receive tutoring.

The problem with these services are: A. the tutors are only vetted by the external service, not by the instructional staff, and so may

	Pre-Exam	Exam 1	Exam 2
Total Students	12	17	14
Students that Dropped	5	5	3
Student Drop Rate	41.67%	29.41%	21.43%

Table 1: Drop Rate by Tutoring Cohort

vary widely in their understanding of the course material, even for courses that do not change much year-to-year. For courses that do change, the tutor may have to learn the material as they are tutoring, diminishing their capacity to help. B. by being external, they have no way of coordinating with other resources, asking for advice from course staff, passing students on to the instructor, suggesting interventions in lab, etc.

Since our internal tutoring team is deeply integrated into the course staff and composed of TAs that are known to understand both the course material in general and the specifics of what is happening in the course, in terms of assignments, labs, etc. – they do not suffer from the lack of utility that the external nature of peer tutoring sometimes causes.

5 Assessment

We present this in two sections: first, we present quantitative data based on how students who we enrolled in the tutoring program did both in the course as a whole and on exams, both before (and compared to the course as a whole) and after.

Next, we present qualitative assessments of the program, drawn from post-tutoring-session surveys that TAs and students completed, an end of semester survey, and from unsolicited comments sent to tutors. While the program was small in scope, we hope that this limited analysis can be useful in showing that it did work as designed, and help others who wish to create similar programs or replicate our design.

5.1 Quantitative

Over the semester, tutoring sessions were administered to a total of 43 students who held a recorded 92 sessions (although this number may be higher as, especially at the beginning, some TAs may have forgotten to report the session).

There are three distinct groups of students we will analyze, referred to as the Pre-Exam, Exam 1 and Exam 2 groups. The course had three mid-term exams and no final: Exam 1 occurred one month into the course, Exam 2 occurred two months into the course, and Exam 3 occurred three months into the course (which was very close to the end of the course, though before the course "drop" deadline).

The Pre-Exam group is made up of students who were identified by the calibration lab described in §3.2, or who were referred directly based on observation by course staff in lab or elsewhere. The Exam 1 group is made up of students who failed the first exam. We took a hybrid approach with the Exam 2 group. After filtering for students who did well on the first exam, we prioritized those who scored the lowest (45%) on the second exam and continued assigning tutors based on availability, eventually reaching students who had scored in the mid 70s on the second exam.

The inaugural goal of this program was to reduce the chances of course failure for students who were identified as "at risk". Since students know all of their exam grades before the drop deadline, this translates to the drop rate acting as an effective failure rate for students. Table 1 shows the drop rates for each of the aforementioned cohorts of students.

While the rates for the Pre-Exam and Exam 1 groups are quite high, we note that these students were all chosen because we believed that without intervention, these students would all fail. Following this hypothesis, our intervention resulted in a majority passing when they may not have otherwise. The Exam 2 group was different: while we were still certainly worried that the downward trajectory between the first and second exam would continue to the third, and thus they may have risked failure, some of them may have been able to scrape by with a passing grade. This generally stronger performance is reflected in the increased pass rate of that cohort.

We also used exam scores to compare student progress in the course, as the exams are cumulative in nature and are held in a controlled environment. This data, shown in Table 2, shows a significant correlation between the administration of tutoring to students and their success in subsequent assessments. In particular, students in both the Exam 1 Cohort and Exam 2 Cohort make a significant increase on the exam following their sessions (Exams 2 and 3 respectively), which is controlled against the class as a whole, indicating that this was not because the subsequent exams were easier. Indeed, the data indicates that Exam 3 was likely slightly more challenging than Exam 2. We do not include data on the Pre-Exam group because we have no objective measure of them before the tutoring intervention.

Additionally, in data collected in a required exit survey, students rated tutor efficacy an average of 9.45 out of 10. While clearly subjective, this leads into our next section – that tutoring can help students feel valued by the course staff.

5.2 Qualitative

While the above quantitative data shows the program had a positive impact, this data alone does not completely represent the impact tutoring has on students who are learning Computer Science. There are increased societal and economic pressures in the current state of the industry to be the "best" which leads to rampant imposter syndrome [5, 6]. This also may discourage students to stay in the field.

Tutoring allows for students to increase their confidence in the course, especially in introductory courses where there are a breadth of skill levels and personal experiences. Additionally, the existence of the tutoring program as an official part of the course, even if not publicly advertised, rather than an external service, helps to normalize it.

Students who responded to an end-of-semester form said: "I felt [tutoring] was a good way to get a one on one understanding of the material especially since a large lecture can be hard to focus", "Tutoring helped me understand and piece together the big topic[s] of [our intro CS course] and clear up any confusions I had", among other positive personal anecdotes. In one of the post-session exit surveys, a tutee captured the positive impact of the proactive design

	Exam 1 → 2 (Exam 1 Cohort)	Exam 1 → 2 (Entire Class)	Exam 2 → 3 (Exam 2 Cohort)	Exam 2 → 3 (Entire Class)
Mean Change	+10.54%	+0.29%	+15.14%	-3.45%
Median Change	+18.79%	0.00%	+13.89%	-2.44%

Table 2: Average Change in Exam Score After Tutoring vs. Control

of the program: "[The tutor] reached out to me and helped me without me seeking anyone. This was the best scenario I could've asked for."

Tutees also often sent unsolicited positive messages to tutors, especially after exams, e.g., "Hey [tutor name] the exam results r out n holy this is my highest score I just wanted to thank you, u r a wizard", and even in cases where they did not perform as well as they wanted to, the individual attention still seemed to help students, e.g., "I got #destroyed by this exam but I wanted to take the time to thank you properly big bro. I know we only had a few sessions together but it really meant the world to me being able to actually get help in that class one on one. Thank you so much for your time this semester I hope you enjoy your break [tutor name]!"

6 Future Work

Our research could benefit from further iteration of the processes we outline above, and long term investigations of student success after this program. Focusing on providing more adaptive and personalized tutoring could also improve the quality of our sessions.

In terms of evaluation, it would be beneficial to track responses on how tutoring impacts the students' motivation, confidence, and engagement in the course from their previous experience without tutoring through short surveys. These surveys can ask for free-response answers or have each student rank their engagement on a 1-5 scale.

Due to the rise of ChatGPT and other LLMs, many students have resorted to AI to complete their assignments, and this can be especially evident when students use content not taught in lectures. Though we highly discourage using LLMs, their usage is unfortunately inevitable. Therefore, another measurement could involve tutees anonymously reporting the effectiveness of help and credibility between LLMs and the tutoring they received.

One way we can improve on personalized tutoring could involve analyzing students' and tutors' personalities and amount of experience to create optimal pairings. Students who are struggling significantly in the course may benefit from pairing with a TA who has more experience by previously TAing the course. A TA who is generally more proactive and has the extra time to reach out consistently may pair best with a student who struggles to ask for help.

7 Reproducibility

Based on the positive results and feedback we received from our prototype tutoring program, we encourage other established universities and institutions to implement similar initiatives, particularly in introductory courses where students often need the most support.

We piloted our program in a first-semester Computer Science course with approximately 600 students and 80 TAs. Around 20% of these TAs were actively involved in the tutoring program, and only about 5% of the students were identified as potentially benefiting from one-on-one support. We believe that institutions with similar student and TA ratios can expect comparable outcomes.

On average, each TA in the tutoring team dedicated only 1 hour per week completing their tutoring responsibilities, though this number varied based on how many students they were currently supporting. The tutoring team leads (the second, third, and fourth authors) spent a similar amount of time weekly, but with a slightly higher average due to the additional time needed for scheduling, checking in with tutors, and evaluating the program's effectiveness. To successfully replicate this program, we recommend designating a few lead TAs to manage coordination and ensure an even distribution of work.

To maintain the on and off-boarding of students and analyze the tutoring program's impact, the tutoring leads used a spreadsheet to record student status, tutor assignments, session content, and academic progress. Formulas we attached to the spreadsheet's cells allowed us to calculate quantitative improvements in assignment and exam scores. We anticipate that similar tools or systems will be necessary for effective implementation at other institutions.

Throughout the semester, the course's lead instructor (the first author) held weekly meetings with the tutoring team leads. These meetings were essential for reviewing the program's quantitative and qualitative performance and making difficult decisions, such as pivoting to support students who had recently fallen behind mid-semester. This consistent and collaborative communication between the instructional team and the tutoring leads was crucial to the program's success.

8 Conclusion

While we have only run this program once, we believe that this structure could easily be replicated in other courses, and the evidence, while limited, seems to suggest that it would be beneficial to do so. The administrative overhead at the instructor level was very low, as it was almost entirely run by undergraduate TAs (the second, third, and fourth author of this report). The total time consumed by all the TAs involved in the program was also, comparatively, quite low: with a staff of 80, we hypothetically had 1200 hours of work per week, but only a maximum of 20 or so was dedicated towards tutoring, or 5%. To get the results we got for 5% of TA time seems absolutely worth it, at least in introductory classes.

Acknowledgments

References

[1] Ildar Akhmetov, Sadaf Ahmed, and Kezziah Ayuno. 2024. How We Manage an Army of Teaching Assistants: Experience Report on Scaling a CS1 Course. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) *(SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 32–38. https://doi.org/10.1145/3626252.3630871

[2] Jens Bennedsen and Michael E. Caspersen. 2019. Failure rates in introductory programming: 12 years later. *ACM Inroads* 10, 2 (April 2019), 30–36. https://doi.org/10.1145/3324888

[3] Shearon Brown and Xiaohong Yuan. 2014. Experiences with retaining computer science students. *Journal of Computing Sciences in Colleges* 29, 5 (2014), 34–41.

[4] Diba Mirza, Phillip T. Conrad, Christian Lloyd, Ziad Matni, and Arthur Gatin. 2019. Undergraduate Teaching Assistants in Computer Science: A Systematic Literature Review. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (Toronto ON, Canada) *(ICER '19)*. Association for Computing Machinery, New York, NY, USA, 31–40. https://doi.org/10.1145/3291279.3339422

[5] Adam Rosenstein, Aishma Raghu, and Leo Porter. 2020. Identifying the Prevalence of the Impostor Phenomenon Among Computer Science Students. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 30–36. https://doi.org/10.1145/3328778.3366815

[6] Angela Zavaleta Bernuy, Anna Ly, Brian Harrington, Michael Liut, Andrew Petersen, Sadia Sharmin, and Lisa Zhang. 2022. Additional Evidence for the Prevalence of the Impostor Phenomenon in Computing. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 654–660. https://doi.org/10.1145/3478431.3499282