(Board 55/Work in Progress) Exploring Office Hour Interactions in a Data Structures and Algorithms Course

Alexander Hicks, Virginia Polytechnic Institute and State University

Alex Hicks (he/him) is a PhD student in the Department of Computer Science at Virginia Tech. He earned a Bachelor of Science in Computer Science and History from the University of Virginia in 2020. His research interests include help-seeking behavior, broadening participation in computer science, and automated feedback systems.

Prof. Cliff Shaffer, Virginia Polytechnic Institute and State University

Dr. Shaffer received his PhD in Computer Science from University of Maryland, College Park in 1986. He is currently Professor of Computer Science at Virginia Tech, where he has been since 1987. He directs the AlgoViz and OpenDSA projects, whose goals resp

WIP: Exploring Office Hour Interactions in a Data Structures and Algorithms Course

Abstract

Large universities often have introductory computing courses with hundreds of students, dozens of TAs, and multiple TAs on duty at the same time. We investigate what occurs during office hour interactions between students and TAs, focusing on a large intermediate data structures course with major programming assignments. We hope to learn what gaps exist in current TA training and open the "black box" of office hours. To collect this data, we are using a self-hosted version of MyDigitalHand, an automated office hours queuing system. We have modified our instance of MyDigitalHand to collect additional data to answer our research questions.

Introduction

Office hours provide the backbone of communication between course staff, specifically teaching assistants, and students. As the number of students in programming courses has increased, universities have responded with three main solutions: 1) increase staff sizes by hiring more Teaching Assistants (TA), 2) use automated feedback systems to take augment instructional staff, and 3) deploy systems designed to support TA staff conducting office hours. Large universities often have introductory computing courses with hundreds of students, dozens of TAs, and multiple TAs on duty at the same time [1], [2]. Managing this army of TAs and ensuring that they are taking the best approaches to helping students can be difficult for instructors, especially since they often get little direct feedback about TA/student interactions.

While we know that students spend time in office hours, especially when they encounter obstacles on their assignments [3], we know surprisingly little about the details of what happens during these interactions. We also do not know exactly what factors make these interactions successful or unsuccessful as defined by the participants.

This study investigates what occurs in office hours in a third year Computer Science course at a large R1 public university. The junior-level Data Structures and Algorithms course includes a major focus on programming assignments, with four projects each with a 3-to-4 week lifecycle. This course sits as a post-CS2 Data Structures and Algorithms course and is a prerequisite for many of the electives offered by the department. Due to the high level of difficulty in the projects, we believe that TAs for this courseoften face challenging questions from students. While students often present their problem as "I need to find the bug", we find that often the "real" problem is some misconception about the assignment, programming language or environment, or their

design. Therefore, TAs should default to pointing students away from a potential code error message, and towards finding the misconception if there is one. In particular, TAs should resist debugging student code.

Since there is not much in the literature about the details of what goes on in student/TA office hours interactions, we see an opportunity here to explore this topic through a data-driven approach. We seek to do this through a combination of data collection (structured responses immediately after an interaction), and post hoc interviews.

To collect data, we used MyDigitalHand [4], an office hour queuing system and research platform. MyDigitalHand allows us to not only create structure and similarity across office hour interactions for a course, it also helps us to get a better idea of what goes on inside of the student-TA interactions themselves. Before joining the queue, students must articulate what they are working on, what problems they have, what they have tried, and their modality. Post-interaction surveys and timing data on the interactions give us hints of what the participants think of as successful or unsuccessful interactions. In the near future, we will follow up on this quantitative data phase with a qualitative data phase consisting of a series of interviews with TAs and students.

Methods

This WIP study takes an explanatory mixed methods approach [5] where we collect quantitative data from MyDigitalHand. Our initial data collection effort contains the interaction times, total questions asked and answered, responses to feedback questions on the topic, and perceptions of success of the interaction. Teaching staff were explicitly asked whether students had misconceptions about what their "real" problem was.

These data allow us to then plan targeted interviews and observations towards interactions identified by TAs and students as unsuccessful to develop a more complete view of office hour interactions at this level.

MyDigitalHand collects two main forms of data: automatically collected data related to the timing or context of the interaction, and after-interaction feedback data requested from both students and TAs on how a session went in their opinion. Automatically collected data consists of time stamps as well as student-provided descriptions of the topic they are seeking help about. After-interaction feedback data contains both student and TA responses to several questions shown in Table 1. These were multiple-choice questions where students and TAs were provided the same answer choices for similar questions [6]. The question of whether or not the problem a student brought to office hours was solved during the session has three possible responses: yes, partially, and no. This question in particular will help us target our interviews for phase two of the study.

The data collected through MDH can indicate some trends of office hours and highlight interesting interactions to investigate more thoroughly. The data provides mechanical context of office hour interactions (who met with who, and for how long), what students say they are working on, how long they talked about it with a TA, and more. We also asked whether an interaction "good" or "bad" without providing guidance to students and TAs on what that meant. Often it is hard to interpret from the data we collect how significant the interaction was. In particular, our current data collection efforts have not allowed us to answer basic questions like:

Student Feedback Questions	TA Feedback Questions
What was your question about?	What was the student asking?
Did you make progress on or solve your problem during your office hour interaction?	What was the student looking for?
	Was the student's problem solved during this interaction?

Table 1: Feedback questions asked of students and TAs with MyDigitalHand

Student ID (anonymized identifier)	1642	1706	1252	1473	1601	1379
TA-Identified Unsuccessful Interaction	3	5	8	2	2	0
Student-Identified Unsuccessful Interaction	9	9	6	6	6	5

Table 2: Number of interactions per student (minimum 5 unsuccessful interactions) that were identified as unsuccessful by the student or the TA involved

Which TAs are doing a good job? Is more time spent per person a good indicator or a bad one? What is a reasonable "success" rate where students indicate whether their question was resolved?

We can identify "interesting" or subpar interactions such as where both participants felt it was unsuccessful, or possibly those where students felt they weren't helped, but TAs felt that they did a good job. Exploring these cases further through interviews will let us see what is happening in these interactions and provide the qualitative data for the mixed methods approach.

Results

Preliminary results from the quantitative data collected allow us to pinpoint which interactions are worth taking a closer look at. We have defined "unsuccessful" interactions as those where the TA or the student indicated that the problem discussed in the interaction was not solved. Table 2 shows the number of unsuccessful interactions per student of those with more than five unsuccessful interactions. Table 3 shows the the number of unsuccessful interactions per TA of those with more than 9 unsuccessful interactions and the percent of unsuccessful interactions per TA of those with at least 10 interactions and at least 10% unsuccessful interactions.

These data provide us with several interview targets, students that had multiple bad experiences with TAs. Talking to these students could expose patterns across the TAs that were not helpful for students, or possibly problems with approaches that certain students take. It is also useful to investigate what these students were seeking from office hours. Without a window into the interaction itself, it is difficult to determine if these students were looking for an answer that the TA was unable to provide due to a lack of understanding, or one that the TA refused to provide due to course policy, for example.

Table 3 shows that some TAs were much more critical of themselves than students were of them, but there were exceptions. TA 1205 felt they were generally doing a decent job of answering

TA ID (anonymized identifier)	1207	1185	1193	1198	1190	1204	1205	1199
TA-Identified Unsuccessful Interaction	32	21	17	17	12	12	12	11
Student-Identified Unsuccessful Interaction	12	9	9	5	3	4	26	9
Percent Unsuccessful Interactions	23.53%	18.42%	15.18%	14.05%	8.76%	15.58%	7.19%	9.48%

Table 3: Number of interactions per TA (minimum 9 unsuccessful interactions) that were identified as unsuccessful by the student or the TA involved and percent of interactions per TA that were identified as unsuccessful by the student or the TA involved

student questions, but the students they interacted with reported a larger number of unsuccessful interactions. With so many interactions identified by students as unhelpful, it could be worth interviewing this TA about how they approached office hours and what strategies they employed to assist students.

There were 21 interactions identified by both the TA and the student involved as being unsuccessful. These interactions are an important subset of those identified in Tables 2 and 3 that should be investigated further. In particular, exploring the actions of the TA, whether they recommended to simply rejoin the queue, seek further clarification themselves from other course resources (or TAs), or moving on to their next student, could spotlight factors that could be addressed in future TA trainings or scaffolding for office hours provided to them.

Discussion

This WIP study focuses on selecting interactions to investigate with interviews based on student and TA feedback about how "successful" the interaction was. From the data we have collected, we plan to expand the criteria for selecting interactions to take into account time spent waiting for and during an interaction. How long a student must wait to see a TA and how long they spend with them once they get matched could be indicators of the quality of an interaction [7].

Time spent on interactions is a concern for this course not just due to potential impact on other students waiting for a GTA. Due to the nature of the course and the difficulty of the projects, students are often asking for help with debugging their program. If a TA accepts that as a reasonable thing to attempt, it can lead to a long interaction time while the TA tries to trace through the program. We generally discourage TAs from taking this approach, but some do this even so.

Another anecdotal concern is that some students will make complaints that they go to TAs who are unable to answer their questions. This is not strongly supported by the actual survey data: while some students report not having a question answered, it does not match the pattern that some students complain about. We hope with better monitoring through the MDH survey process, as well as better understanding about student/TA interactions in general, that we will have a better way to recognize and address these situations. Ideally we would be able to identify whether some TAs really are unhelpful or whether some students have poor help-seeking behaviors.

Universities have started creating training programs for TAs [8], [9]. However, these programs are often not well-formed or standardized. These programs can be as structured as a full course, as

short as a one hour seminar, or completely non-existent [10]. Even when they do exist, these programs often have gaps in what training they provide for TAs or focus on completing trainings required by the university instead of instruction on how to better interact with students [10].

Developing a better idea of what occurs in office hour interactions will allow us to more accurately design systems to scaffold and provide TAs with the necessary information to help students, or to better prepare students to attend office hours productively.

Limitations

This work is limited by the ability of the researchers and course staff to require use of the data collection platform by students and TAs. It is important to note that we are not merely surveying students and TAs for voluntary feedback. Data collection is embedded within a useful office hours queuing tool. This provides incentive for the users to participate, especially during peak hours. But since we are new to adopting MyDigitalHand within our department, it can take time to establish a "new normal" of routine use.

A second limitation is relying on the ability of students and TAs to self-identify how good or bad an interaction is without additional context. Some TAs might have no experience to compare to, or may only be comparing to their own experiences as students in other courses. For GTAs, this might be at other universities and in other countries.

Future Work

Now that students and TAs have been identified, the next steps are to interview those selected and attempt to match their feedback about what happens inside of office hours to markers in the quantitative data to allow more effective measurements of office hours to be taken at scale.

This work is the first step towards defining and gathering better metrics for TA performance. Current methods of evaluating TA performance are often subjective reviews from instructors that might not ever see the TA actually perform in office hours [11], [10]. Investigating a data-driven approach that accurately measures TA performance based on objective measures of their interactions with students can provide insights that allow for better TA training systems. Without knowing where office hours are unsuccessful, it is difficult to design useful interventions to better prepare TAs. With the data collected from this work, we have the opportunity to visualize where TAs could benefit from additional scaffolding in office hours.

References

[1] M. Ball, J. Hsia, H. Pon-Barry, A. DeOrio, and A. Blank, "Teaching TAs To Teach: Strategies for TA Training," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, (Portland OR USA), pp. 477–478, ACM, Feb. 2020.

[2] E. McDonald, G. Arevalo, S. Ahmed, I. Akhmetov, and C. Demmans Epp, "Managing TAs at Scale: Investigating the Experiences of Teaching Assistants in Introductory Computer Science," in *Proceedings of the Tenth ACM Conference on Learning @ Scale*, (Copenhagen Denmark), pp. 120–131, ACM, July 2023.

[3] S.-H. Ko and K. Stephens-Martinez, "What Drives Students to Office Hours: Individual Differences and Similarities," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, (Toronto ON Canada), pp. 959–965, ACM, Mar. 2023.

[4] A. J. Smith, K. E. Boyer, J. Forbes, S. Heckman, and K. Mayer-Patel, "My Digital Hand: A Tool for Scaling Up One-to-One Peer Teaching in Support of Computer Science Learning," in *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, (Seattle Washington USA), pp. 549–554, ACM, Mar. 2017.

[5] "Designing and Conducting Mixed Methods Research — SAGE Publications Inc." https://us.sagepub.com/en-us/nam/designing-and-conducting-mixed-methods-research/book241842.

[6] L. Perlmutter, J. Salac, and A. J. Ko, ""A field where you will be accepted": Belonging in student and TA interactions in post-secondary CS education," in *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1*, (Chicago IL USA), pp. 356–370, ACM, Aug. 2023.

[7] Z. Gao, S. Heckman, and C. Lynch, "Who Uses Office Hours?: A Comparison of In-Person and Virtual Office Hours Utilization," in *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, (Providence RI USA), pp. 300–306, ACM, Feb. 2022.

[8] F. Muzny and M. D. Shah, "Teaching Assistant Training: An Adjustable Curriculum for Computing Disciplines," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, (Toronto ON Canada), pp. 430–436, ACM, Mar. 2023.

[9] A. Zaman, A. Cook, V. Phan, and A. Windsor, "A Practical Strategy for Training Graduate CS Teaching Assistants to Provide Effective Feedback," in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, (Turku Finland), pp. 285–291, ACM, June 2023.

[10] D. Mirza, P. T. Conrad, C. Lloyd, Z. Matni, and A. Gatin, "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), pp. 31–40, ACM, July 2019.

[11] E. Patitsas and P. Belleville, "What can we learn from quantitative teaching assistant evaluations?," in *Proceedings of the Seventeenth Western Canadian Conference on Computing Education*, (Vancouver British Columbia Canada), pp. 36–40, ACM, May 2012.