What Drives Students to Office Hours: Individual Differences and Similarities

Shao-Heng Ko
shaoheng.ko@duke.edu
Duke University
Durham, North Carolina, USA

Kristin Stephens-Martinez
ksm@cs.duke.edu
Duke University
Durham, North Carolina, USA

ABSTRACT

Undergraduate teaching assistants (UTAs) ofice hours are an approachable way for students to get help, but little is known about why and for what do the students choose to attend ofice hours. We sought to understand what kind of help the students believe they need by analyzing the problem-solving step students self-reported when joining the ofice hours queue app. We used the UPIC frame-work to aggregate course specific problem-solving steps to enable comparing between seven data sets from a CS1 and a data science course across four semesters. We then compared the class-level and student-level phase distributions to understand the differences be-tween the two courses and the two levels in the courses. We found most students have a "primary phase" where a majority of their interactions fall, and there are significant individual differences in their phase distributions. Moreover, we did not find either students' demographics or the context of their first visits to significantly impact their individual differences in the phase distributions, sug-gesting students may have fixed beliefs on how to approach ofice hours. Finally, a strong majority of interactions happen within 3 days of the deadline, such that the UPIC distribution for those days looks like the class-level phase distribution.

CCS CONCEPTS

• Social and professional topics → Computing education; CS1.

KEYWORDS

CS1, data science, ofice hours, undergraduate TAs, problem-solving process, help-seeking behavior

ACM Reference Format:
Shao-Heng Ko and Kristin Stephens-Martinez. 2023. What Drives Students to Ofice Hours: Individual Differences and Similarities. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023), March 15–18, 2023, Toronto, ON, Canada. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3545945.3569777

1 INTRODUCTION

Instrumental help seeking is positively related with student performance [4]. This kind of help-seeking is where a student is taught how to solve the problem, rather than given the answer. Well-trained undergraduate teaching assistants (UTAs) can give students such help in ofice hours. Students can find this source of help "relatable and approachable" [10] as they provide less formal interactions than instructors or graduate TAs. However, little is understood on what issues drive students to seek help in UTA ofice hours and the commonalities across the ofice hour interactions. By understanding what they seek help on, we can provide more targeted training to UTAs. Moreover, such information can help a teacher decide if the reason a student seeks help is healthy for their learning and inform how they motivate the students to go to ofice hours.

We took the UPIC framework [14] to better understand each individual student's reason for attending UTA ofice hours. UPIC unifies different problem-solving processes by dividing them into common phases. We labeled each interaction with a UPIC phase using the student's self-reported problem-solving step. Our goal was to understand what phase the student both believed they needed help on and decided to seek help on it. In addition, by comparing a student's phase between visits, we can understand how strong a particular phase motivates a student to attend ofice hours.

Our research questions are:

(1) How representative is the class-level overall phase distribution compared to an individual student's phase distribution?
(2) Do frequent ofice hours users have a primary phase that account for a majority of their visits? If so, how strong is it?
(3) What influences the individual differences in primary phases?
(4) Do the due dates influence the phase distributions?

We collected data from a CS1 and a data science (DS) course across four 15-week semesters starting with Fall 2020 for a total of seven data sets (4 from the CS1 and 3 from the DS) using the My Digital Hand [13] ofice hours app. In this work, we report general information about our data sets similar to related work [6, 13].

We found that students are each unique in what UPIC phase drives them to go to ofice hours. As such, the student-level UPIC phase distributions are not similar to the class-level distribution, meaning the latter should not be used to assume what phase is driving a particular student to ofice hours. In addition, most students have a majority of their phases falling in one primary phase. We did not find this individual difference to be driven by the student's demographics or the first visit's phase.

When dividing the time before a deadline into 3-day windows, we found that around half of the students only utilized ofice hours within 3 days of the deadlines. In addition, we did not find significant differences among the time windows for the class-level interaction distributions because a strong majority of interactions happened within 0-3 days to the deadlines.

SIGCSE 2023, March 15–18, 2023, Toronto, ON, Canada

ACM ISBN 978-1-4503-9431-4/23/03…$15.00
https://doi.org/10.1145/3545945.3569777

2 RELATED WORK

2.1 Ofice Hour Tool and Data Analysis

My Digital Hand (MDH) [13] by Smith et al. is an ofice hours queue management tool that facilitates data collection of one-to-one interaction between the students and the TAs. The tool supports customized and open-ended pre- and post-interaction surveys. Smith et al. also analyzed the wait time, interaction duration, and average number of ofice hour visits per student, finding variances in data from three universities across two semesters.

Several subsequent papers also analyzed ofice hours data collected via MDH [5–7, 14, 15]. Gao et al. [6] compared the utilization of in-person and virtual ofice hours from classes before and after the Covid-19 pandemic. The UPIC framework [14] divides the problem-solving process into four phases: (1) Understand the problem, (2) create a Plan, (3) Implement the plan, and (4) verify Correctness/debug. We used the same data sets as this work that only analyzed the class-level interaction distribution of UPIC phases.

Using the Design Recipe [3] paradigm on program design, Ren et al. [12] also analyzed students' self-assessed needs in ofice hours, as well as the alignment between the students' and the TAs' assessments. Departing from the conventional one-on-one ofice hours model, the work of Campbell and Craig [1] investigated the usage, wait time, and service time of an alternative drop-in help center that catered to students in multiple courses.

These works focus on characterizing a baseline of what is happening in the ofice hours. We seek to understand students' motivations for going to ofice hours, which may shed light on how to improve them and influence attendance.

2.2 Student Help-Seeking Behavior

Going to ofice hours is one of many kinds of student help-seeking behavior. Among the eight not necessarily sequential stages of the help-seeking process outlined by Karabenick and Dembo [9], the action of going to ofice hours belongs to the solicit help stage for a student. However, collecting their current problem-solving phase sheds light on other stages (such as determine whether there is a problem) that led to the decision to solicit help.

The meta-analysis by Fong et al. [4] summarizes the current literature on academic help-seeking behavior of college students. Two key factors of student help-seeking behavior identified in their literature review are the goals and sources of help. Common types of goals in students' help-seeking behavior include instrumental (emphasizing on the "process" of acquiring problem-solving skills), executive (focusing on the "outcome" of getting things done), and avoidant help-seeking (referring to a denial of seeking help even if it is determined necessary). They found a positive relationship between instrumental help-seeking and academic performance, while other types of behaviors appeared to be detrimental. On the other hand, the source-of-help spectrum spans from formal (e.g., from instructor or course staff) to informal help (e.g., from peers or web resources). While computing students tend to seek help from informal sources first [2], only formal help was found to have a significant correlation with achievement in the meta-analysis study.

Ofice hours run by well-trained UTAs provides a source of instrumental help, while at the same time sits at an intermediate position on the formal-informal spectrum: they are simultaneously formal (as teaching staff) as well as informal (as peers or near-peers). UTAs have been in CS classes for a long time [11], but little work reports specifically on their effect on students through ofice hours [10]. Our work seeks to start unpacking what is happening in ofice hours and why, especially whether they help retain the positive effect of formal instructions while also reduces avoidant behavior because they are less formal than the primary teacher.

Table 1: Summary of class format and structure. Instructors with an asterisk in co-teaching arrangements are primary.

Course-	semester	Instruction format	Ofice Hours	Instructor	Autograder usage
CS1	Fa20	Remote	Remote	A*, B	Yes
	Sp21	Remote	Remote	A, B	Yes
	Fa21	Hybrid	Hybrid	B	Yes
	Sp22	Hybrid	Hybrid	A	Yes
DS	Sp21	Remote	Remote	C*, D	No
	Fa21	Hybrid	Remote	D	Yes
	Sp22	Hybrid	Hybrid	D	Yes

3 METHODS

3.1 Class Setting and Participants

The data consists of UTA ofice hour interactions from classes at Duke University, a medium-size, research-oriented, private university. We collected seven data sets: four consecutive 15-week semester offerings of a CS1 course from Fall 2020 (Fa20) to Spring 2022 (Sp22), and three consecutive offerings of an intermediate-level data science (DS) course for Sp21 - Sp22. Both courses are taught in Python. The CS1 focuses on learning programming without assuming prior experience, whereas the DS emphasizes using libraries and data analysis techniques and has CS2 as a prerequisite. Due to the Covid-19 pandemic, the classes were offered in either a fully remote format over Zoom [16] or a hybrid format over Zoom and in-person. Each course had two primary instructors. Each offering was taught by one or both instructors. The format and structure of all classes and ofice hours are summarized in Table 1.

As summarized in Table 2, the consent rate, utilization rate of ofice hours, demographics, and prior experience vary between the classes. Due to small numbers and privacy requirements, all students that reported non-binary genders were marked as no response. Moreover, we replaced all race counts less than 5 with an asterisk, and omitted the columns with only asterisks. We also collected student graduation year; however, we do not report that information due to the lack of space. Note the number of serviced students (went to ofice hours) is noticeably smaller than the class population, limiting the generalizability of what we report. Moreover, due to reporting on only consenting students, sometimes the demographics are biased, such as for CS1 Sp21 in which only 54.7% men gave consent compared to 75.6% women.

3.2 Data Collection

We collected interaction data from MDH [13], where each interaction takes place between a student and an UTA. Before the interaction, the student was asked what part of the problem-solving

Table 2: Student demographics. Total is the number of students enrolled in that class. The consent rate is the percent of consenting students. The service rate is the number of student served over the number of consenting students. The frequent rate is the number of students that used MDH at least three times over the number of student served. For all demographic subcategories, the number outside the parenthesis is the number of consenting students in that category, and the number inside is the number of consenting students served in that category. Black includes African American, and 2+ stands for multiracial. Values lower than 5 are replaced with an asterisk. The threshold between less and more prior experience differs by course. Not all students answered all questions.

Course-Semester		Total	Consenting (consent rate)	Served (service rate)	≥ 3 visits (freq. rate)	Gender		Race				Hispanic	Prior Experience	
						Men	Women	White	Asian	Black	2+		Less	More
CS1	Fa20	198	152 (76.8%)	68 (44.7%)	37 (53.6%)	75 (32)	73 (35)	74 (35)	42 (13)	12 (8)	13 (5)	19 (10)	100 (47)	51 (21)
	Sp21	216	157 (72.7%)	90 (57.3%)	58 (64.4%)	47 (16)	93 (63)	65 (31)	50 (30)	8 (6)	11 (9)	15 (6)	127 (70)	14 (9)
	Fa21	241	177 (73.4%)	89 (50.3%)	50 (56.1%)	64 (28)	94 (52)	80 (40)	48 (22)	15 (8)	12 (6)	22 (14)	124 (68)	37 (12)
	Sp22	221	152 (68.8%)	77 (50.7%)	34 (44.2%)	48 (19)	76 (44)	63 (31)	39 (19)	* (*)	15 (9)	22 (10)	97 (46)	27 (17)
DS	Sp21	217	181 (83.4%)	38 (21.0%)	14 (35.9%)	79 (11)	93 (27)	76 (15)	69 (17)	12 (*)	12 (*)	16 (*)	61 (16)	86 (18)
	Fa21	198	144 (72.7%)	48 (33.3%)	23 (47.9%)	78 (17)	62 (30)	61 (19)	55 (21)	9 (5)	10 (*)	10 (*)	51 (14)	64 (18)
	Sp22	209	145 (69.4%)	48 (33.1%)	26 (54.2%)	84 (22)	55 (21)	59 (20)	63 (15)	8 (*)	5 (*)	11 (7)	35 (14)	75 (20)

Table 3: Summary of interactions. Numbers in parenthesis is the proportion over raw data count.

Course-Semester		# raw	# A (consenting)	# B (valid)	# C (valid UPIC)	# D (days to deadline)
CS1	Fa20	828	549 (66.3%)	489 (59.1%)	398 (48.1%)	341 (41.2%)
	Sp21	931	808 (86.8%)	759 (81.5%)	650 (69.8%)	555 (59.6%)
	Fa21	1006	740 (73.6%)	653 (64.9%)	515 (51.2%)	439 (43.6%)
	Sp22	574	458 (79.8%)	406 (70.7%)	289 (50.3%)	224 (39.0%)
DS	Sp21	216	162 (75.0%)	145 (67.1%)	136 (63.0%)	N/A
	Fa21	269	216 (80.3%)	195 (72.5%)	173 (64.3%)	
	Sp22	354	225 (63.6%)	202 (57.1%)	189 (53.4%)	

process they needed help with. In the CS1 data sets, they also reported what they were working on at the assignment level. At the end of the interaction, students indicated whether they made progress and rated their experience. After the interaction, the UTA also answered all of the (pre and post) questions the student answered about the interaction. We collected all responses, as well as the timestamps when a student joined the queue and when an interaction started and ended. For this work, we focused mainly on students' responses on what they needed help with, and their post-interaction feedback, detailed below.

Question on what the student needed help with. The CS1 class used Hilton et al. 's [8] seven steps, while DS used their own problem-solving steps.[1] We chose to follow the UPIC framework [14] and the mappings provided there to bucket the steps into UPIC phases. This enables an apples to apples comparison between the different problem-solving steps in the CS1 and DS courses. For CS1 Sp21-Sp22, this question was multiple choice, whereas for all other data sets, students could select all options that applied. Among the data sets where the student could choose multiple options that covered multiple phases, 16.9%-28.3% of the interactions corresponded to multiple phases, and only 3.2%-5.8% involved more than two. Whenever the student did so, we used the earliest phase in the UPIC framework. Rounding "up" biases our results towards earlier in the problem-solving process and assumes that students are more likely earlier than where they think they are.

Post-interaction feedback. For both courses, the students reported whether they made progress. For CS1, the students also rated their satisfaction on a instructor designed Likert scale of 1 to 4.

Table 3 summarizes the different subsets of interactions we used for analysis. Data set A includes all the consenting students (63.6%-86.8%). Data set B is a subset of A where the waiting time is at most four hours and the interaction duration is between one to sixty minutes, and thereby represent valid MDH interactions (57.1%-81.5% of the raw data).[2] Data set C is the subset of interactions in B where the student indicated at least one UPIC phase (48.1%-69.8% of the raw data). This set filters out interactions unrelated to problem solving (e.g., about technical issues). Data set D is the subset of C where the student indicated an assignment they were working on, and the interaction happened before the deadline of the assignment.[3] This subset excludes interactions for other course contents. In this work, we use subset A for research questions only involving general office hour usage, B for research questions that involve student experiences but not the UPIC framework, C for research questions that involve breaking down UPIC phases, and D for research questions that involve analyzing the number of days between interactions and deadlines.

4 RESULTS

4.1 General Information

Figures 1a, 1b, and 1c depict the empirical cumulative distribution functions (ECDF) of the wait time, the interaction duration, and the number of interactions per student for each dataset. The wait time distributions behave similarly to that in Smith et al. [13] and

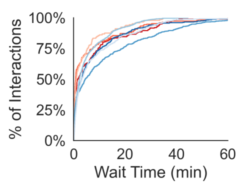
[1] Each class also had other miscellaneous options such as "Tech Issue".

[2] Records that fall outside these ranges may represent technical errors (e.g., UTA forgetting to close interactions) or an interaction that did not actually happen (e.g., student was away from keyboard when it was their turn). We used the same interaction duration thresholds as Gao et al. [6], and the waiting time upper bound reflects that no office hour session was longer than four hours in the courses.

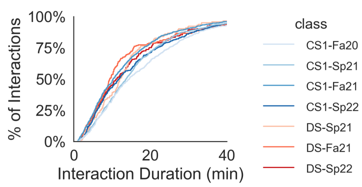
[3] Interactions that happened after the corresponding deadline could be the student selecting a wrong assignment.

Table 4: Statistics of interactions. All time is in minutes.

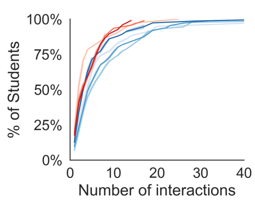
Course-Semester		Wait time		Interaction duration		Visits per student	
		Mean	Median	Mean	Median	Mean	Median
CS1	Fa20	8.59	3.15	17.76	14.42	7.19	4
	Sp21	5.70	2.22	16.12	13.60	8.43	5
	Fa21	12.59	4.53	13.50	10.25	7.42	4
	Sp22	8.07	3.04	15.50	11.65	5.41	3
DS	Sp21	4.62	1.02	14.14	11.93	3.84	2
	Fa21	6.39	0.72	13.63	9.35	4.24	3
	Sp22	7.78	1.53	14.68	11.07	4.21	3



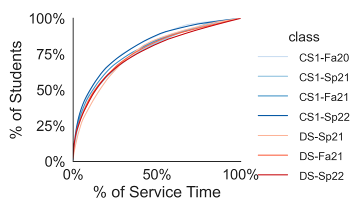
(a) ECDF of wait time.



(b) ECDF of interaction duration.



(c) ECDF of number of interactions per student.



(d) Proportion of students vs. their total service time used.

Figure 1: General Information. To improve legibility, we cut off the long tails by omitting less than 10% of the interactions.

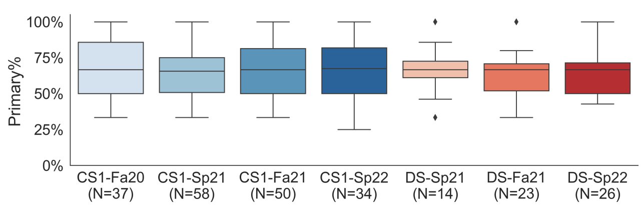


Figure 2: Box plots of primary% of students. Students with only 1 or 2 interactions are excluded.

Gao et al. [6]: they are usually under 20 minutes, but can go over an hour when the demand is high, resulting in a long tail. As such, none of the data are normally distributed. To better convey the central tendency, we report both the medians and means in Table 4.

Figure 1d shows the percentage of the total service time (x-axis) used by a fixed proportion of students (y-axis). We plotted this figure by first sorting students by their total amount of service time (with least first), then summing up the cumulative time using this order. While Smith et al. [13] found that the 5% most heavy users took up 50% of total service time, our classes had 10-20% of most prolific students take up 50% of the resources.

4.2 Students' individual phase tendencies and primary phases

We next investigate the students' reasons for attending ofice hours, borrowing the UPIC framework [14]. Instead of looking at the overall class-level distribution of UPIC phases in the classes in the original work [14], we seek to understand: (1) what each individual student seeks help on, and (2) how similar are the students' individual interaction UPIC distributions to the class-level distribution.

To this end, we first calculated the class-level proportions of interactions for each UPIC phase per data set to create a four element vector.[4] Next, we calculated the same vector for each student. We then took the difference between the class-level and student vectors and measure its 1-norm, i.e., the sum of absolute values of its elements. This can be interpreted as the "distance" between the student's individual distribution and the class-level distribution. Therefore, if students are fully homogeneous in their UPIC tendencies, then we would expect a value of 0 for every student. On the other hand, if every student only goes to ofice hours in one UPIC phase, then we would expect four different "spikes" at those distances, each corresponding to one UPIC phase.

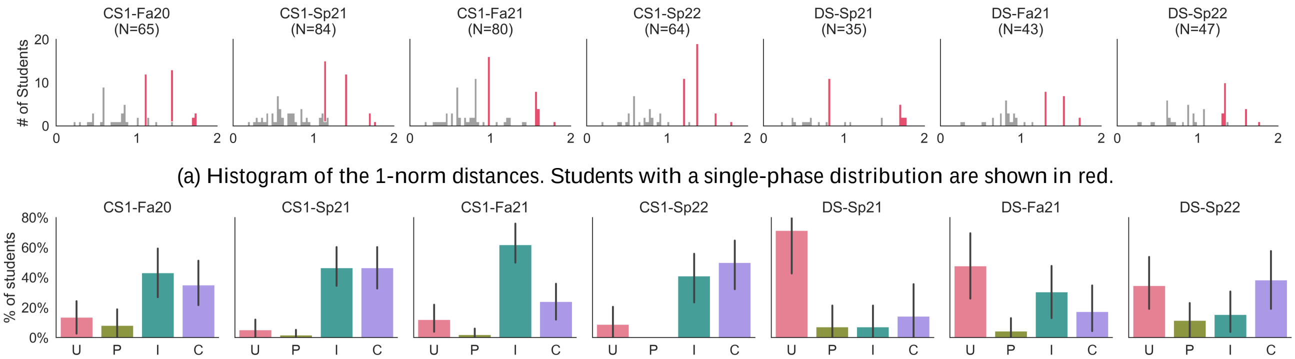
Figure 3a shows the empirical distribution of the 1-norm distance in each data set. The observed distance values suggest the students are more heterogeneous than homogeneous in their UPIC phase tendencies, or in other words, the class-level distribution provides little information about a given student's distribution.

We then ask: do students have a specific UPIC phase for a majority of their interactions? And if so, what is the distribution of this primary phase? We defined a student's most frequent UPIC phase as their primary phase, as well as primary% the proportion of their interactions in the primary phase.

Figure 2 shows the empirical distribution of the primary% of the students. Here we only show the primary phases of the frequent students (at least 3 interactions), which accounts for 35.9%-64.4% of the served students (see Table 2). This is to avoid an excess of 100% and 50% primary% values for the students who only used ofice hours once or twice. As shown, the median primary% values are between 60% and 70%, and all the lower quartiles are at or above 50%, suggesting that the students' primary phases indeed account for a majority of their interactions. Figure 3b depicts the empirical distribution of the primary phases of the frequent students. Every UPIC phase has students with that phase as their primary phase, except for Plan in CS1-Sp22. In sum, our data suggests that a significant part of the class-level experience is in fact driven by the individual differences of students' approaches.

To investigate what causes the students to differ in their individual UPIC phase tendencies, we first examine whether there exists an "anchoring effect", where the students continue to use ofice hours based on their first interaction. To investigate this we compared two distributions using the Kolmogorov-Smirnov test. The first distribution was each student's proportion of interactions for their first interactions' UPIC phase. The second distribution was the proportion from a random visit. In none of our 100 experiments was the test statistically significant. In other words, we found no evidence that the students are "primed" by their first interaction. This finding suggests that students have fixed beliefs of how they

[4] Please see [14] for visualizations of these distributions.



(a) Histogram of the 1-norm distances. Students with a single-phase distribution are shown in red.

(b) Proportion of students with their primary UPIC phases. Students with only 1 or 2 total visits are excluded.

Figure 3: Visualizations of students' individual UPIC phase tendencies.

approach ofice hours, and they are formed before the start of the courses. Alternatively, students may have different primary needs in the problem-solving process. We also examined if the students' primary phases are correlated with their demographics and found no significant correlation between the primary phases of students and their gender, race, or prior coding experiences.

4.3 Do due dates impact phase distributions?

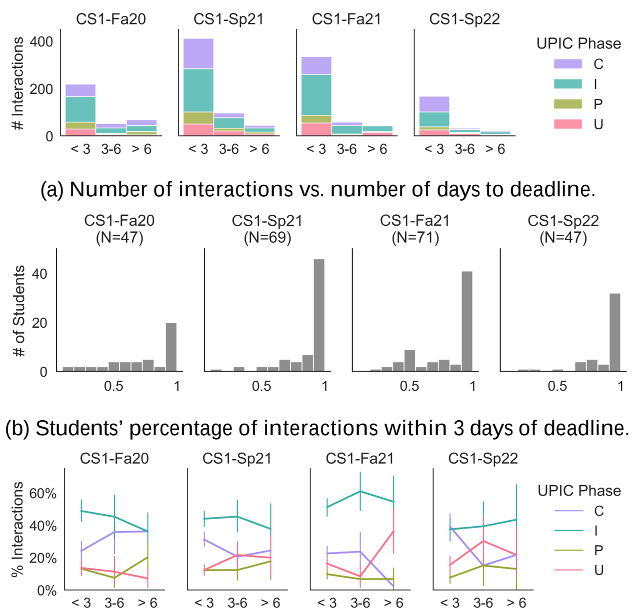
To further understand students' help-seeking behavior, we shift the focus to investigate whether the relative "nearness" to the assignment due dates influences the UPIC phase distributions on either the class-level or the individual-level. To measure the nearness, we calculated the number of days between each interaction and the due date of the assignment that the student was working on. Note that we can only infer this information for CS1 as students did not specify their assignment in DS.

Figure 4a plots the number of UPIC interactions for each 3-day window (i.e., interactions less than 3 days before the deadline, 3-6 days before, and so on). Consistently, a strong majority of interactions happened within 3 days from the deadlines of the assignments, verifying the anecdotal belief that students are driven by deadlines.

On the individual level, Figure 4b plots the histogram of students' percentage of UPIC interactions within 3 days of the deadlines. A significant portion of students only attended ofice hours within 3 days of deadlines (40.4%-68.1%).[5] In the Sp21 and Sp22 semesters, such students were a clear majority.

To understand whether there is a shift in the students' UPIC phase tendencies in different windows, we plot (in Figure 4c) the proportions of interactions in each UPIC phase for each 3-day window. Unsurprisingly, Implement is the most prominent phase with Understand and Plan at the bottom. Surprisingly, Correctness is not consistently highest during the 0-3 day window and decreasing as there is more time before the deadline. We suspect the Correctness phase trend could still be true for individual students, but as students may have various timelines of completing the assignment (in terms of when they start and finish), we could not identify the expected phase shift in the aggregated overall distribution. Ideally, we would like to: (1) identify the students who start early (and

[5]Without filtering out non-UPIC interactions, this proportion was 40.4%-64.8%.



(a) Number of interactions vs. number of days to deadline.

(b) Students' percentage of interactions within 3 days of deadline.

(c) Proportion of UPIC interactions vs. number of days to deadline.

Figure 4: Analysis of ofice hours utilization and UPIC phase distributions in different stages of the course.

regularly utilize ofice hours) and those who do not, and study their respective UPIC phase distributions; and (2) break down the set of interactions to the assignment level, as anecdotal evidence suggests students have different needs for different types of assignments. However, the number of students/interactions is often too small, preventing us from digging further and looking at those levels.

5 LIMITATIONS

There are multiple limitations in our data sets. First, our consent rate fluctuated per data set, and we know that the distribution of demographics is not always representative of the class as a whole (see Table 2). If different demographic groups use ofice hours differently, our methodology cannot easily account for this. Moreover, anecdotally we know the fall and spring semester CS1 demographics are often markedly different, adding another layer of complexity. Second, the data set size limited any findings' strength when aggregated or filtered (see Table 3). The number of interactions in the CS1-Sp22 data set is lower than the other three CS1 classes, and the number of students in the DS was smaller than the CS1 in terms of the raw count and how many went to ofice hours. Finally, the data sets straddle different instruction settings and contexts. Across Table 1's four variables describing the data sets, none of the data sets have the same setting nor context. There may be teaching or modality effects, and we were not always able to capture the interaction's modality.

Another limitation is in the data collection methodology. The students' self-reported data only reflects what the students thought was happening when they filled out the pre-/post-interaction survey, instead of what actually happened during the interaction. Therefore, the self-report's accuracy hinges on the students' metacognitive skills. Furthermore, the pre-interaction survey reflects the reason that student decided to join the queue; they may make progress while waiting for a TA. Consequently, there may be significant noise in linking students' post-interaction responses with the UPIC phase of the interaction. One potential remedy is validating the students' data by the TA's responses to the same questions, which occur after the interactions. However, this requires TA consent and potentially losing more data. The accuracy of TA's responses may also suffer from lack of attention (due to insuficient training and/or dealing with a busy queue). Therefore, we chose not to include the TA's responses in this work.

6 DISCUSSION/FUTURE WORK

Students are unique snowflakes. The main insight from our analysis on students' UPIC phase distributions is that students need very different kinds of help in their own problem-solving process, not only in different classes but also within the same class. As such, the class-level distributions do not accurately represent the reasons that drive individual students to ofice hours. In contrast, past ofice hour usages at the individual level can help the course staff understand/infer (and accommodate) a student's need. Given the diversity in students' individual phase distributions, there may be value having TAs specialize in different problem-solving process phases, and match the students' needs with the TAs' specialities.

Causes of individual UPIC differences. Within a class, our analysis did not identify any factor that correlates with the students' individual differences in their phase distributions. None of the students' first visit, gender, race, and prior experience appeared to be significantly relevant. We suspect that students have their own fixed ideas about how to approach or use ofice hours. These fixed opinions may have formed during their first exposure to ofice hours in the curriculum, instead of the first visit for the class, and may also be impacted by other factors such as peer influence. In addition, autograders (as well as how students use them) may factor into their phase distributions [14]. More effort, potentially including studies on students' help-seeking behavior across different classes, is needed to identify the causes of their individual differences.

Interplay between student demographics and ofice hour experiences. We also investigated whether students from different demographics have different behavior/experiences in their usage of ofice hours. We did not find any statistically significant evidence that suggests students with different gender/race/prior programming experience approach ofice hours differently in their usage rate, visit frequency, or phase distribution. Similarly, we did not see any difference in their wait time, interaction duration, or the feedback outcomes. However, given our limitations, we do not think demographic differences can be entirely ruled out.

What about being driven away from ofice hours? While the UPIC framework provides a lens to understand what kinds of help drives students to ofice hours, a perhaps equally important question is what drives them away from ofice hours. Anecdotally, students may be driven away by negative experiences, such as no progress or poor treatment. To this end, we analyzed the students' feedback on their first visit and last visit. We hypothesized the first visit may represent the student's "first impression", whereas the last visit may hint at a negative experience that influenced the student to stop coming. However, for the first visit, we did not find any feedback options with a significant correlation to the number of times the student came back. Similarly, we did not find any evidence suggesting a difference in their experiences in their last visit compared to an average visit; if anything, the average last visit appeared to be more satisfactory (and make more progress) than the average visit. However, this approach is limited: we could merely observe that the student did not come back, instead of distinguishing between those that did and did not need to come back. A larger-scaled study with more specific focus on the students' willingness and need to use the ofice hours regardless of if they did is likely needed to investigate this direction.

7 CONCLUSION

We sought to understand what drives students to go to UTA ofice hours. By analyzing data collected from a CS1 and a data science course across a two-year span via the UPIC framework, we identified significant individual differences in what motivates them to attend ofice hours. More specifically, most students have a primary phase in the problem-solving process in which they go to ofice hours for a majority of their visits. We found neither demographics nor the first-time visits to significantly impact such differences, which hints students could have fixed beliefs of how to use ofice hours. We also observed that students most utilize ofice hours near assignment deadlines, but otherwise did not find gender, race, or prior experience to be significant factors in their ofice hours usage; however, we do not think demographic differences can be entirely ruled out. Our work is a primary attempt to unpack not only what is happening in ofice hours but why, in hope of sparking more in-depth analysis on the raised questions.

ACKNOWLEDGMENTS

We thank Sadhana Suryadevara and Manith Luthria for their early analysis of the data. This material is supported by the National Science Foundation grant #1934965.

REFERENCES

[1] Jennifer Campbell and Michelle Craig. 2018. Drop-In Help Centres: An Alternative to Ofice Hours. In WCCCE '18: Proceedings of the 23rd Western Canadian Conference on Computing Education. ACM, 9:1–9:6. https://doi.org/10.1145/3209635.3209642

[2] Augie Doebling and Ayaan M. Kazerouni. 2021. Patterns of Academic Help-Seeking in Undergraduate Computing Students. In Koli Calling '21: 21st Koli Calling International Conference on Computing Education Research. ACM, 13:1–13:10. https://doi.org/10.1145/3488042.3488052

[3] Matthias Felleisen, Robert Bruce Findler, Matthew Flatt, and Shriram Krishnamurthi. 2018. How to Design Programs: An Introduction to Programming and Computing. MIT Press. http://htdp.org/

[4] Carlton J. Fong, Cassandra Gonzales, Christie Hill-Troglin Cox, and Holly B. Shinn. 2021. Academic help-seeking and achievement of postsecondary students: A meta-analytic investigation. Journal of Educational Psychology (2021), Advance online publication. https://doi.org/10.1037/edu0000725

[5] Zhikai Gao, Bradley Erickson, Yiqiao Xu, Collin Lynch, Sarah Heckman, and Tiffany Barnes. 2022. Admitting you have a problem is the first step: Modeling when and why students seek help in programming assignments.. In EDM 2022: Proceedings of the 15th International Conference on Educational Data Mining. International Educational Data Mining Society, 508–514.

[6] Zhikai Gao, Sarah Heckman, and Collin Lynch. 2022. Who Uses Ofice Hours?: A Comparison of In-Person and Virtual Ofice Hours Utilization. In SIGCSE 2022: Proceedings of the 53rd ACM Technical Symposium on Computer Science Education, Vol. 1. ACM, 300–306. https://doi.org/10.1145/3478431.3499334

[7] Zhikai Gao, Collin F. Lynch, Sarah Heckman, and Tiffany Barnes. 2021. Automatically classifying student help requests: a multi-year analysis. In EDM 2021: Proceedings of the 14th International Conference on Educational Data Mining. International Educational Data Mining Society, 81–92.

[8] Andrew D. Hilton, Genevieve M. Lipp, and Susan H. Rodger. 2019. Translation from Problem to Code in Seven Steps. In CompEd '19: Proceedings of the ACM Conference on Global Computing Education. ACM, 78–84. https://doi.org/10.1145/3300115.3309508

[9] Stuart A. Karabenick and Myron H. Dembo. 2011. Understanding and facilitating self-regulated help seeking. New Directions for Teaching and Learning 2011 (2011), 33–43. Issue 126. https://doi.org/10.1002/tl.442

[10] Diba Mirza, Phillip T. Conrad, Christian Lloyd, Ziad Matni, and Arthur Gatin. 2019. Undergraduate Teaching Assistants in Computer Science: A Systematic Literature Review. In ICER '19: Proceedings of the 2019 ACM Conference on International Computing Education Research. ACM, 31–40. https://doi.org/10.1145/3291279.3339422

[11] Stuart Reges, John McGrory, and Jeff Smith. 1988. The Effective Use of Undergraduates to Staff Large Introductory CS Courses. ACM SIGCSE Bulletin 20, 1 (1988), 22–25. https://doi.org/10.1145/52964.52971

[12] Yanyan Ren, Shriram Krishnamurthi, and Kathi Fisler. 2019. What Help Do Students Seek in TA Ofice Hours?. In ICER '19: Proceedings of the 2019 ACM Conference on International Computing Education Research. ACM, 41–49. https://doi.org/10.1145/3291279.3339418

[13] Aaron J. Smith, Kristy Elizabeth Boyer, Jeffrey Forbes, Sarah Heckman, and Ketan Mayer-Patel. 2017. My Digital Hand: A Tool for Scaling Up One-to-One Peer Teaching in Support of Computer Science Learning. In SIGCSE '17: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. ACM, 549–554. https://doi.org/10.1145/3017680.3017800

[14] Sadhana Suryadevara and Kristin Stephens-Martinez. 2022. UPIC a Problem-Solving Framework: Understand, Plan, Implement, and Correctness/Debugging. In ICER '22: Proceedings of the 2022 ACM Conference on International Computing Education Research, Vol. 2. ACM, 50–51. https://doi.org/10.1145/3501709.3544286

[15] Matthew Zahn, Lina Battestilli, and Sarah Heckman. 2022. Describing Academic Help Seeking Patterns in Introductory Computer Science Courses. In 2022 ASEE Annual Conference & Exposition. https://peer.asee.org/41526

[16] Zoom. 2022. https://zoom.us/.