The Trees in the Forest: Characterizing Computing Students' Individual Help-Seeking Approaches

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

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

ABSTRACT

Background and Context. Academic help-seeking is vital to post-secondary computing students' effective learning. However, most empirical works in this domain study students' help resource selection and utilization by aggregating the entire student body as a whole. Moreover, existing theoretical frameworks often implicitly assume that whether/how much a student seeks help from a specific resource only depends on context (the type of help needed and the properties of the resources), not the individual student.

Objectives. To address the gap, we seek to investigate individual computing students' help-seeking approaches by analyzing what help-seeking characteristics are *individual-driven* (and thus stay consistent for the same student across different course contexts) and what are *context-driven.*

Method. We analyzed $N = 597$ students' survey responses on their help resource utilization as well as their actual help-seeking records across 6 courses. We examined relations between *individual students'* frequency-based help usage metrics, type-of-help requested in office/consulting hours, self-reported *order of ideal help resource usage*, and their collaboration inclination in small-scale sections.

Findings. We found that students' frequency-based help metrics and their order of ideal help resource usage stays relatively consistent across different course contexts, and thus may be treated as part of students' individual help-seeking approaches. On the other hand, the type of help students seek in office/consulting hours and how much they collaborate with peers in small sections do *not* seem to stay consistent across different contexts and thus might be deemed more context-driven than individual-driven.

Implications. Our findings reveal that part of students' help-seeking characteristics is individual-driven. This opens up a possibility for institutions to track students' help-seeking records in early/introductory courses, so that some preliminary understanding of students can be acquired before they enter downstream courses. Our insights may also help instructors identify which part of students' help-seeking behavior are more likely to be influenced by their course context and design.

ICER '24 Vol. 1, August 13–15, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0475-8/24/08.
https://doi.org/10.1145/3632620.3671099

CCS CONCEPTS

• **Social and professional topics → Computing education**.

KEYWORDS

Computing Education, Help-Seeking

ACM Reference Format:
Shao-Heng Ko and Kristin Stephens-Martinez. 2024. The Trees in the Forest: Characterizing Computing Students' Individual Help-Seeking Approaches. In *ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3632620.3671099

1 INTRODUCTION

Academic help-seeking is both a metacognitive behavior and a self-regulated learning (SRL) strategy [28, 29, 40, 42] adopted by students to support their learning. The positive effects help-seeking can bring on students' academic achievement, especially with *instrumental* help-seeking approaches that emphasize mastery/acquiring skills, are widely known and studied [16]. Existing educational theories on help-seeking, such as Karabenick's eight help-seeking stages [29] and Makara and Karabenick's expectancy-value model and four dimensions of help resources [42], have motivated many subsequent works on factors in students' *help resource selection and utilization* [12, 20, 48, 49, 51, 55, 60].

However, most of these works implicitly model students' help-seeking decision/resource selection as *one* standalone process that depends almost solely on context. Some of these works study properties of the current help need, such as type of help desired [42] and effort to phrase the current help need [51]; others study properties of the help resources, such as accessibility/availability [12, 42, 48, 51, 60], perceived usefulness [12, 42, 60], and synchronicity/timeliness [20, 51]. In contrast, the effects of personal *approaches* that differ from individual to individual are less emphasized in these models, despite recent explorative works showing substantial individual differences in students' help-seeking behavior in both computing [33, 52, 63] and other [20, 55] contexts.

One possible reason for this gap in existing literature lies in the sparsity of help-seeking records. Thanks to the booming growth of enrollment and Ed-tech platforms in computing-related fields in the recent years, computing educators now have access to many kinds of data that track students' learning. However, *not all students seek help*, and only a tiny portion of them help-seek frequently enough for fine-grained analyses. Furthermore, help-seekers do not always utilize class-affiliated help resources that enable data collection, and their interactions with *informal* resources such as classmates, static online resources, or general large language models (LLMs) are usually not captured. As a result, *student-specific* help-seeking data remains relatively harder to collect than other kinds of data such as students' incremental programming trace in controlled IDEs.

As a result, most recent works on students' help-seeking records focus on one or two specific help resources in one specific instructional context, while *individual* students' behavior across different instructional contexts is less studied. This gap has created a dissonance between what the field has studied (mostly treating the entire class/cohort of students as a whole) and what educators ultimately care about (we want to provide *each* student adequate and personalized help that they need, or motivate *every* student that needs help to seek it). Motivated by these gaps, our work seeks to study computing students' *individual* help-seeking approaches and records in multiple instructional contexts, using both self-reported and actual behavioral data. Our research question is:

- (RQ) What student help-seeking characteristics are individual-driven (part of their individual approach), and what others are context-driven?

To tackle this research question, we surveyed students on their help resource utilization (11 offerings across 6 courses) and collected students' help-seeking records (20 offerings across the same 6 courses) in formal (course-affiliated) resources. We then filtered for students that appeared in at least two courses in our data ($N = 597$ after filtering), and examined the correlations and associations between their help-seeking characteristics in different courses.

Our analyses reveal that some aspects of a student's help-seeking approach seem individual-driven while others are more driven by the course context. From actual help-seeking records, our analyses revealed that students' *frequency-based help metrics* (such as number of threads a student initiated on the class forum) in different courses are strongly correlated, while their distributions of type of help requested in office/consulting hours are not. Therefore, the former is more likely part of students' individual help-seeking approaches/habits, while the latter might be more heavily influenced by the course context. From the survey data, we also found students' self-reported *order of ideal help resource usage* to stay rather consistent across different courses, justifying it being a part of students' help-seeking approach. On the other hand, how much students collaborate with others in small-section discussion/lab sections does not seem to stay consistent across courses.

Although limited by the scope of our data collection, our work serves as an explorational attempt to characterize computing students' individual help-seeking approaches. On the practical side, our findings reveal that part of students' help-seeking characteristics is individual-driven and might preserve across different contexts, opening up the possibility for institutions to track students' help-seeking records along the entire curriculum so as to better support each type of student in downstream courses. Through differentiating *individual-driven* and *context-driven* characteristics, our insights may also help instructors identify parts of students' help-seeking behavior that are directly influenced by the design of their course help ecosystem, and make progress towards helping students find the approach that best supports their learning. For research purposes, our categorization of characteristics can serve as a primer for clustering students with similar characteristics together and subsequently studying each "cluster/persona" via more in-depth qualitative research methods.

2 RELATED WORKS

2.1 Help-seeking theories

Help-seeking as a self-regulated learning strategy. Academic help-seeking is a metacognitive behavior deeply intertwined with self-regulated learning (SRL) [28, 29, 40, 42]. Karabenick and Dembo [29] break down the academic help-seeking process into eight *stages* (that are not always sequential): (1) determine whether there is a problem; (2) determine whether help is needed/wanted; (3) decide whether to seek help; (4) decide on the type of help (goal); (5) decide on whom to ask; (6) solicit help; (7) obtain help; and (8) process the help received. As pointed out by Stites *et al.* [55], phase 2 (identifying the need for help) corresponds to the *forethought* phase in Zimmerman's SRL model [64], whereas phase 8 (processing the help and determining whether further help is needed) aligns with the *self-reflection* phase.

Type of help. Early studies on help-seeking [26, 27, 30] focused on investigating phase 3-4 of the 8-stages model (decide whether to seek help and if so, what type of help to seek) by examining factors such as self-efficacy and self-esteem. A taxonomy for categorizing the type of help is *instrumental* versus *executive/expedient* [16, 34, 35]: the former focuses on acquiring *skill/mastery* and thus usually reduces the need for future help, while the latter prioritizes the *outcome* and thus only has a short-term effect (work gets done) but not necessarily improves learning. Fong *et al.*'s meta-analysis [16] found only instrumental help but not executive/expedient help to be positively related to academic performance.

Resource selection/utilization. Although the original 8-stage model phrased phase 5 as *decide on whom to ask*, the variety of help resources students can access has far outgrown the initial wording. As such, Makara and Karabenick [42] contextualize students' *perceptions* of help resources by the *role* (formality of the resource), the *relationship* (between students and the resource), the *channel* (that help gets delivered), and the *adaptability* (of the help resource to different individual's needs). Based on these dimensions, they refined phase 5 of the 8-stage model by a *resource selection expectancy-value model*: the likelihood a student utilizes a resource relies on both the resource's *accessibility/availability* and how much the student's desired type of help matches what the resource provides.

Inspired by Makara and Karabenick's model, many subsequent works investigated students' resource selection process [12, 20, 48, 49, 51, 55, 60], revealing an ever-growing list of factors that influence students' help-seeking behavior such as *formality* of the resource [49], *timeliness of response* [20], *trust in the resource* [48], and *synchronicity* and *effort to phrase a problem* [51]. Wirtz *et al.* [60] and Doebling and Kazerouni [12] subsequently take the model one step forward: they study students' *self-reported frequency of use* and *perceived usefulness* of all available help resources, and identify an *order of usage* progression from *less social, less formal, and more accessible* resources to *more social, more formal, and less accessible* resources. Based on their efforts, Hou *et al.* [25] recently compared students' self-reported frequency of using large language models (LLMs) and perceived usefulness against existing help resources. While these works provide deep insights into students' help resource selection and utilization, they are mostly based on students' self-reported usage patterns instead of actual behavioral data.

2.2 Resources-Specific Works

Social help resources. Thanks to data collection in Ed-tech platforms, many existing works have reported and analyzed computing students' *social* help-seeking behavior in each platform: course-specific discussion forums [8, 52, 54, 57, 59], public discussion forums [13], video platforms [4], study groups [46], in-person [53] and online [18] course-specific office/consulting hours, and non-course-specific academic help centers [5]. These works usually focus on documenting the course-aggregated usage statistics such as usage rates (what proportion of students used the specific resource, and its relation with different student groups) [8, 18, 33, 58, 59, 62], usage frequencies (how many times a student used the specific resource on average, again across different demographic/identity groups) [33, 52, 54, 57, 58, 62], and resource-specific statistics such as *average wait time* and *interaction length* in office/consulting hours [18, 19, 33]. Some also categorize the kind of help students seek from these resources [19, 33, 50, 56, 59] and/or the overall *quality of help-seeking attempts* (e.g., whether students provide sufficient context when asking a question) [17, 19, 59]. Finally, several works analyzing *timings* of help-seeking behavior [4, 17, 33] found students' help-seeking behavior peak immediately before assignment deadlines, and the type of help students seek differ between peak and non-peak time.

Non-social help resources. The help computing students receive is not limited to social help from humans. Automated feedback in learning management systems (LMS) [9] and autograders [32] have been widely used to scale instruction and provide adaptive [43] or formative [21, 22] feedback on students' works. Due to their highly synchronous and accessible nature, another line of works study (and attempt to mitigate) students' *over-reliance* on automatic feedback tools [1, 38], even before the transforming emergence of large language models (LLMs) sparked many discussions on how (should) LLMs fit in computing/programming classrooms [3, 36, 37, 41]. Several works in this wave benchmark LLMs' responses to novice programmers' help requests [23] and novice programmers' interactions with generative AI tools [31, 47]. Hou *et al.* [25] recently found a bimodal distribution of usage of LLMs as a help resource: students that have adopted LLMs rely on it almost daily, whereas many students never use it at all.

Resource interaction. Although most works focus on investigating one resource (or two resources *separately*) at a time, several works have examined the *interaction* between different resources. Deorio and Keefer [10] report that providing more transparent automated feedback (e.g., making test cases visible) reduces the demand and waiting time in office/consulting hours. Two other works [2, 61] studied how the existence of autograders influence the kind of help students seek in introductory programming courses. Liu *et al.* [39] incorporated LLMs into their course-affiliated ecosystem and found this to have reduced students' social help needs.

2.3 Synthesis and our contribution

Despite all the efforts, most of our understanding on computing students' help-seeking behavior (on both the theoretical and empirical sides) implicitly treat the student body as a whole and characterize what happens in *one* instructional context. Outside of computing education, there has been efforts on modeling students' individual help-seeking characteristics so as to find a few canonical help-seeker "profiles" [15], "clusters" [7], or "patterns" [14], but those works draw from a snapshot of each student's help-seeking characteristics, which may include context-driven factors. Deviating from them, our work attempts to characterize help-seeking behavior/approach from an *individual* student's perspective and study what is/is not stable for individual students *across course contexts*.

3 PARTICIPANTS AND DATA COLLECTION

Table 1 summarizes our data collection. Our data is collected at Duke University, a medium-size, research-oriented, private university in the Southeastern United States with 15-week semesters. The data spans 6 courses: introduction to programming (CS1), data structures and algorithms (DSA), data science (Data), discrete math (DM), database systems (DB), and algorithm design/analysis (Algo). The data is collected from Fall 2020 (Fa20) to Fall 2023 (Fa23) while the number of offerings for each course varies (see Table 1). Throughout this paper, we refer to the six courses above as *courses*, and single offerings of a course as *classes* (20 in total).

Parts of the data were collected during the Covid-19 pandemic with abnormal instructional contexts. Specifically, all synchronous class meetings were exclusively online over Zoom [65] up to Sp21, then shifted to a hybrid format in Fa21 where both in-person and Zoom modalities were available. Data remained hybrid throughout the rest of our scope, while CS1 switched back to in-person in Sp22. We report the participant demographics in Table 4 in the Appendix.

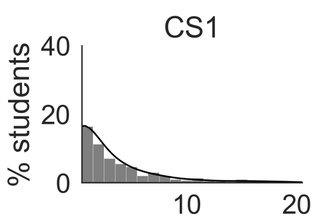
3.1 Frequency-based help metrics

Class forum. All classes used Piazza [45] for the class forum up to Sp21 and Ed discussions [11] since Fa21. Both data sources captured the timestamps of all thread initiations and comments/responses, total number of days each student was active on the forum, and how many posts/comments/responses each student read. When aggregating how many threads/comments/responses each student *contributed* (i.e., wrote), we exclude all comments/responses to threads initiated by a staff member. These posts are mostly announcements instead of help-seeking attempts. Due to different classes having different policies on regulating the use of private/anonymous posts on their class forums, we did not analyze the number of private/anonymous posts for each student in our analyses.

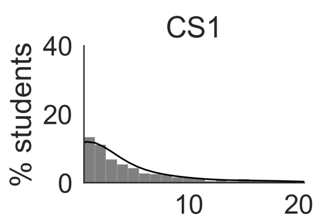
Consulting hours. Throughout this paper, we use *consulting hours* to refer to what past works also called *office hours* or *tutoring hours*: a designated time/place where students get helped by a class staff member (an instructor, a graduate TA, or an undergraduate TA) on *one-on-one* basis. For all classes in our data, most of their regular consulting hours were offered in the evenings, and most of help interactions were carried out by undergraduate TAs. All classes collected the timestamps of interactions between students and TAs, but only Data-Sp23 and Data-Fa23 collected student-instructor interactions. We excluded the interactions where the wait time (the time the student stayed in the queue) exceeded four hours or the interaction duration was not between 1-60 minutes; such interactions were mainly due to human errors.

Table 1: Summary of participants and data collection. An × in a cell represents that data was collected from that class. For the Order survey, × represents the survey was administered once at the midpoint of the semester, and a 2 represents the survey was administered twice, at the midpoint and end of the semester. CS1-Fa23 is omitted: being the entry course, no students took it with another course simultaneously. There is no data for DSA from Fa21-Fa22 as consent was not collected. Class forum usage data includes timestamps of each post/comment/response, total number of posts/comments/responses read, and total number of days active on forum for each student. Consulting hours usage data includes timestamps of each interaction, for which the type of help requested was only collected in some classes (due to platform differences).

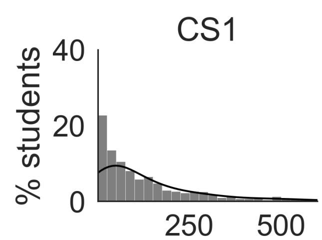
Course	Semester	Participants		Class forum	Consulting hours		Questionnaires	
		Total	Consenting	Usage	Usage	Type of help	Order	Collab
CS1	Fa20	198	152 (76.8%)	×	×	×		
	Sp21	216	157 (72.7%)	×	×	×		×
	Fa21	241	177 (73.4%)	×	×	×		×
	Sp22	221	152 (68.8%)	×	×	×		×
	Fa22	262	183 (69.9%)	×	×	×		×
	Sp23	218	163 (74.8%)	×	×	×	×	×
DSA	Sp21	276	217 (78.6%)	×	×	×		×
	Sp23	331	238 (71.9%)	×	×		2	×
	Fa23	390	272 (69.7%)	×	×		2	×
Data	Sp21	217	181 (83.4%)	×	×	×		
	Fa21	198	144 (72.7%)	×	×	×		
	Sp22	209	145 (69.4%)	×	×	×		
	Fa22	208	152 (73.1%)	×	×	×	×	
	Sp23	234	160 (68.4%)	×	×	×	×	
	Fa23	82	66 (80.5%)	×	×	×	×	
DM	Sp23	128	94 (73.4%)	×	×		×	×
	Fa23	120	94 (78.3%)	×	×		2	×
DB	Fa23	330	183 (55.5%)	×	×	×	×	
Algo	Sp23	298	202 (67.8%)	×	×		2	×
	Fa23	160	123 (76.9%)	×	×		2	×



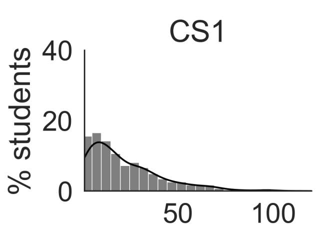
(a) # of Class Forum Threads Initiated.



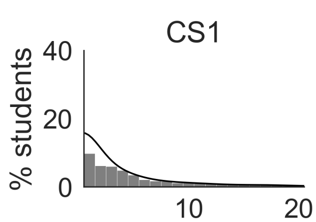
(b) # of Class Forum Contributions (Threads/Comments/Answers).



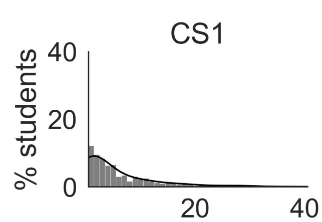
(c) # of Total Views on Class Forums.



(d) # of Days Active on Class Forums.



(e) # of Consulting Hours Interactions.



(f) # of Social Help Requests (sum of Parts 1a and 1e).

Figure 1: Histograms and KDEs for frequency-based help metrics across all CS1 offerings.

Frequency-based help metrics. Figure 1 plots the histograms for each frequency-based help metric for CS1, where the bar heights represent the percentage of students across all CS1 offerings (see Table 1) with that many help requests/events. Similar to what has been reported by past works [18, 33, 50], the distributions of *active* [63] help requests exhibit an extremely right-skewed long-tail behavior with a mode of zero, including number of forum threads initiated (Figure 1a), number of forum threads/comments/answers contributed (Figure 1b), number of consulting hours interactions (Figure 1e), and number of social help requests (the sum of the number of forum threads initiated and the number of consulting hours interactions, plotted in Figure 1f). On the other hand, distributions of *passive* help requests are still right-skewed but do not have modes at zero. This includes number of total views on the forums (Figure 1c) and number of days active on the forums (Figure 1d).

We report the complete set of *course-level* histograms (i.e., a histogram that aggregates across all offerings of a *course* for each metric) in Figure 6 in the Appendix. The help demands vary from course to course as part of their instructional contexts, with the most salient observable differences being the consulting hours interactions distributions (See Figure 6e in the Appendix). This combined with the fact that none of the six metrics are close to being normally-distributed motivates normalizing each metric by *percentiles* before our analyses (see Section 4.1).

3.2 Type of help in consulting hours

In addition to the frequency-based help metrics, CS1, Data, and DB collected the *type of help requested* in each consulting hours interaction via a pre-interaction survey in the MyDigitalHand queueing app [44]. Students were required to fill out this survey before submitting each help request in MyDigitalHand. On the other hand, DSA, DM, and Algo used an internal queueing platform where the type of help was not collected.

The MyDigitalHand pre-interaction survey captures where the students' help needs were in terms of their problem-solving processes. CS1 used Hilton *et al.*'s [24] seven steps for the options, while Data and DB used another problem-solving framework detailed in Table 5 in the Appendix. We follow the *Understand, Plan, Implement, and Correctness* (UPIC) framework [56] and the mappings provided there to bucket the problem-solving steps in each class into UPIC phases to allow for inter-class comparisons. The type of help question was multiple choice in CS1 Sp21-Sp23 and select-all-that-apply in all other classes. Note that the type of help question is asked once per interaction, and thus is fundamentally different than the per-student surveys that we outline below.

3.3 Per-student surveys

Order of help resource usage. In Data-Fa22 and all classes since Sp23, we surveyed students on their *order of help resource usage* (the Order survey hereon). The survey was administered at (or slightly later than) the midpoint of each semester. In some of the classes (see Table 1), the survey was also repeated at the end of the semester.

The survey asks all students to *group and rank* all available help resources in the class in their ideal *order of resource usage.* The survey's list of available help resources depended on each class' context. More specifically, we asked the students to "put the action(s)/resource(s) that they utilize first if/when they need help (*assuming available*) in group 1", and then "put the action(s)/resource(s) that they turn to *when the first group of action(s)/resource(s) are unavailable or not helpful enough* in group 2", and so on. Students could use up to 3 groups and could omit resources that they did not use. Please see Table 6 in the Appendix for a detailed table of available help resources in each class.

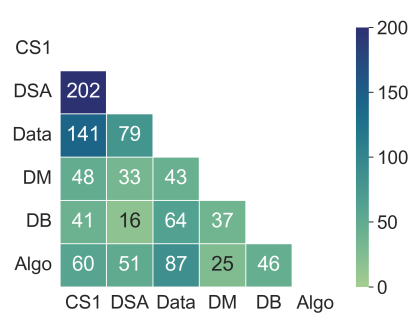
Note that the Order survey asks for students' *ideal* order of resource usage (which is agnostic to the availability/accessibility of all resources) rather than their *typical* order of resource usage (which is highly influenced by availability of each resource in the course's help ecosystem). Therefore, this survey is fundamentally different from asking students to report their *frequencies* of using each of the resources [12, 55, 60] (see Section 4.5 for more discussions).

Collaboration. All courses except Data had discussion/lab sections in which students were *expected* but not *enforced* to work with peers on course material. These classes administered a Collab survey (along with the Order survey, whenever the latter was present). In the Collab survey, students were asked to report their sentiments on the claim "*I work collaboratively with other students in discussion/lab*" from a 5-options Likert question with options from "*Strongly disagree*" to "*Strongly agree*".

3.4 Participants filtering

To analyze students' help-seeking characteristics across course contexts, we filter for students that appear in two or more courses/classes in our data (no students repeated any course). This leaves 597 students in our sample that appeared in 2-5 courses/classes. Specifically, 464/100/30/3 students appeared in 2/3/4/5 courses/classes respectively (see Figure 2a), while no students appeared in 6 courses/classes.

(a) Histogram of # of courses (classes) students appeared in.



(b) Heatmap of course intersections.

Figure 2: Students participating in multiple datasets.

Table 2: Spearman's ρ test results between students' raw/normalized frequency-based help metrics in two different classes. All p-values are Bonferroni-adjusted and are significant at the 0.05 significance level after adjustments. 95% confidence intervals of ρ are obtained via Fisher's z-transform and are raw (unadjusted).

Metric	(Corresponding Figure)	Raw metrics ρ	Raw metrics 95% CI of ρ	Raw metrics Adjusted p	Normalized percentiles ρ	Normalized percentiles 95% CI of ρ	Normalized percentiles Adjusted p
Number of class forum threads initiated	(Figure 6a)	0.47	[0.43, 0.59]	5.9e-35	0.44	[0.39, 0.55]	1.2e-30
Number of class forum contributions	(Figure 6b)	0.47	[0.42, 0.59]	1.4e-34	0.44	[0.39, 0.55]	6.3e-31
Number of total views on class forums	(Figure 6c)	0.47	[0.43, 0.59]	2.8e-35	0.58	[0.58, 0.74]	2.4e-56
Number of days active on class forums	(Figure 6d)	0.52	[0.49, 0.65]	3.4e-43	0.56	[0.56, 0.72]	3.3e-52
Number of consulting hours interactions	(Figure 6e)	0.41	[0.36, 0.52]	6.9e-27	0.40	[0.35, 0.51]	8.6e-26
Number of social help requests	(Figure 6f)	0.49	[0.45, 0.61]	7.8e-38	0.49	[0.46, 0.62]	2.1e-38

Figure 2b demonstrates, for each pair of courses, the number of students that appeared in both courses in our data (regardless of which classes). Note that students that participated in more than two courses/classes appear more than once in this heatmap, and thus the sum of all nonzero cells exceeds the total number of students in our analyses.

4 RESULTS

4.1 Frequency-based help metrics: *frequent help-seekers remain frequent help-seekers*

We examined frequency-based help metrics and found them to be more individual-driven than context-driven. For all 597 students that participated in multiple classes in our data (see Figure 2), we obtained their help platform usage metrics in the (chronologically) first and last classes they took in our data. For tie-breaking among classes in the same semester, we treated the lower-divisional class (with the smaller course number at the institution) as the "earlier" class. By picking each student's two courses that are farthest away from each other from either the chronological or the curriculum perspective, we are likely to get the "most distant" snapshots of the students' help-seeking approach. The more distinct instructional contexts we examine, the more likely that any effect that we may observe is individual-driven.[1]

For each frequency-based help metric listed in Section 3.1, we ran a Spearman's ρ test between students' metric values for their *earliest* classes and that for their *latest* classes. We use the non-parametric Spearman's ρ test because none of the metrics are normally distributed. To further control *course*-specific contexts (such as the varying levels of help demand related to course contents) and *class*-specific contexts (such as total number of assignments that heavily influences help-seeking frequencies), we normalized each metric by percentiles among all students in the respective *class*.

Table 2 shows, for each frequency-based help metric, the Spearman's ρ test results for both the raw metric and the normalized percentiles. Regardless of being normalized or not, every metric exhibits a strong ordinal correlation effect (ρ spans 0.41-0.52 for raw metrics and 0.40-0.58 for normalized percentiles). In other words, a student that seeks help more frequently according to any one of these metrics in their earlier class (compared to their peers in the class, if normalized) tends to also seek the same form of help more frequently in their later class, and vice versa.

Collectively, these results hint *individual social help-seeking frequencies relate heavily to students' personal approaches.*

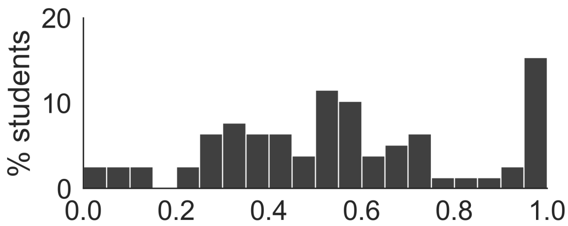
4.2 Students' type of help requested in consulting hours *does NOT seem to stay consistent across courses*

We found the type of help students request in consulting hours to be more context-driven than individual-driven. Our analyses for type of help in consulting hours are based on a smaller subset of 78 students due to the sparsity of help records. More specifically, among the 597 students that participated in multiple classes in our data, only 277 (46.4%) of them participated in multiple classes in which the type of help requested in consulting hours interactions were collected (see Table 1 and Figure 2b). Furthermore, only 90 students (32.5% of the 277) utilized consulting hours in two or more classes where the class collected the type of help requested. Finally, only 78 students (86.7% of the 90) had at least one UPIC-captured interaction in two or more classes.[2] Our analyses throughout this subsection are thus based on these 78 students.

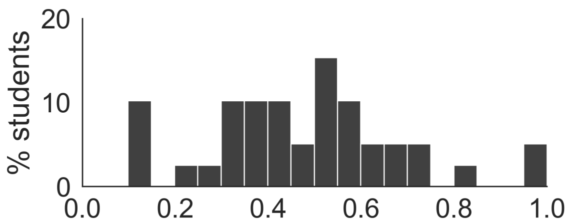
For each student and each pair of their classes, we measured the difference between the student's consulting hours interactions for those two classes (in terms of problem-solving phases) using the Cityblock distance between normalized distributions of UPIC-captured consulting hours interactions. Intuitively (and informally), for each pair of their classes, we compare the two corresponding normalized distributions and obtain how much probability weight is "shifted" between the two distributions. For example, consider a student who has two interactions (one Understand and one Plan) in class A and four interactions (all in Understand) in class B. The two normalized distributions can be represented as $\langle \frac{1}{2}, \frac{1}{2}, 0, 0 \rangle$ and

[1]We acknowledge here that a more robust approach would be to analyze the intersection of each pair of courses/classes separately, so that we do not rely on this assumption about instructional contexts, but our data is too sparse for this approach. However, note that most of our students (464/597, see Figure 2a) only appear in two courses in our data, and thus the choice of using the two farthest-away courses only really affects 133/597=22.3% of the pairs. As a baseline, we also ran the same analysis after *randomly shuffling* each student's records, which completely disregards any chronological/curriculum consideration. All frequency-based metrics still remain strongly correlated under this setting.

[2]In other words, 12 of the 90 students only utilized consulting hours for non-problem-solving-related reasons (e.g., seeking technical help on installing software/configuring environments) in at least one of their classes.



(a) All students (N = 78).



(b) Heavy users (3+ interactions per class, N = 39).

Figure 3: Distribution of mean Cityblock distance. Data has $N = 78$ students; 68/9/1 students had at least one consulting hours interaction captured by the UPIC framework in 2/3/4 classes, respectively. 39 out of the 78 students had at least 3 UPIC-captured interactions in multiple classes.

$\langle 1, 0, 0, 0 \rangle$ respectively for the two courses. As such, the Cityblock distance between these two vectors is $\frac{1}{2}$, as a probability weight of $\frac{1}{2}$ "shifted" from the Understand phase to the Plan phase from class A to class B. This metric has a range of 0 to 1: a distance of 0 happens when the two distributions are identical, and a distance of 1 happens when the two distributions are completely disjoint.

For students with more than two classes's worth of UPIC-captured consulting hours records, we took the mean of the Cityblock distances over *all pairs* of their classes. As such, each student's mean Cityblock distance is still between 0 and 1, capturing (in aggregate) how similar/different their consulting hours usage records across all of their classes are, in terms of problem-solving phases.

Figure 3a shows the distribution of all $N = 78$ students' mean Cityblock distances. While our previous work on type of help sought in consulting hours [33] found that most consulting hours users have an individual "primary UPIC phase" that the majority of their records fall in, our analysis here reveals that *these primary phases, and students' distributions in general, do not seem to stay consistent from course to course.* Indeed, a student with the same "primary UPIC phase" throughout all of their classes can never have a mean Cityblock distance of 0.5 or higher,[3] and yet the median student in Figure 3a had a mean Cityblock distance of 0.51.

Of particular note, a significant portion (12/78 = 15.4%) of the students have a mean Cityblock distance of 1.0. However, these are mainly students with very sparse interaction records in some or all of their classes: if a student used consulting hours only once in each of their two classes, and the two interactions are not of the same UPIC phase, they would easily have a mean Cityblock distance of 1.0. Such a result speaks more volume about the sparsity of their records than whether their type of help sought in consulting hours remain consistent across courses.

Therefore, following the earlier work [33], we discarded all student records with less than three UPIC-captured interactions in a class, keeping only the heavy consulting hours users in the dataset. Figure 3b plots the distribution for the remaining $N = 39$ (50.0%) students who had 3+ UPIC-captured interactions in multiple classes. As shown in the figure, after further filtering away sparse records, the mean Cityblock distances for heavy users are still quite high (median = 0.48), showing substantial distributional difference across students' own consulting hours usage records.

In all, our results suggest *type of help requested in consulting hours more depend on course context and less on individual approaches.* In other words, *the distribution of type of help requested should not be considered part of a student's individual help-seeking approach.*

We also repeated the same analysis using another slightly different metric (the mean Hellinger distance [6]), yielding similar results. Please see Appendix B for the formal mathematical definitions of all metrics used (including for Cityblock) as well as the results for the mean Hellinger distance.

4.3 Order of help resource usage *may be a key part of student's help-seeking approach*

We found the order of students' ideal help resource usage to be more individual-driven than context-driven. Due to the Order survey not administered in earlier semesters (see Table 1), only 260 among the 597 (43.6%) students that participated in multiple classes in our data answered the Order survey in multiple classes. Recall that each time a student answered the Order survey, they provided an *order function* from the set of all available help resources in that class's ecosystem (see Table 6 in the Appendix) to the *used first*, *used second*, *used third*, and *never used (NA)* groups in the survey.

We thus need more metrics to quantify the disagreement of two order functions. Table 3 lists all four metrics we define–*Normalized Undirected Disagreement* (NUD), *Normalized Directed Disagreement* (NDD), *Normalized Pairwise Swap* (NPS), and *Normalized Toggled Resources* (NTR)–between two order functions as well as a running example using five resources.

These metrics play different roles in capturing how a pair of order functions are "different from each other". The *Normalized Undirected Disagreement* (NUD) is the most intuitive and captures how much the two functions "disagree" with each other. However, it also *overestimates* the disagreement, as it inherently captures the potential difference in students' help need in different contexts; if a student has the same ideal order of help resource usage in two classes, but simply does not need as much help in class B as in class A, they might put *every* resource in a later group in class B than in class A, which would result in a NUD of 1.0 (100%). The NDD metric captures this phenomenon as a baseline that quantifies the overestimation in NUD: if all the resource movements are simply due to the student "moving the entire function forward/backward", we can expect their NDD to be equal to or near their NUD.

Unlike NUD, NPS attempts to conservatively measure the disagreement between two functions in a *pairwise* fashion: the relative order of two help resources being strictly swapped between the

[3]This is because the probability weight attributed to their primary phase would be larger than 0.5 in all of their empirical distributions.

Table 3: Definition and running example on metrics for order functions. The definitions are based on two general order functions $f_1 : \mathcal{R}_1 \to \{1, 2, 3, 4\}$ and $f_2 : \mathcal{R}_2 \to \{1, 2, 3, 4\}$, where $\mathcal{R}_1$ and $\mathcal{R}_2$ are the set of available help resources in the two classes' help ecosystems, and $1, 2, 3, 4$ represent the *used first*, *used second*, *used third*, and *never used* groups in the Order survey, respectively. For the running example, consider the common help context of 5 help resources $\mathcal{R} = \{r_a, r_b, r_c, r_d, r_e\}$, where a student indicated $g_1(r_a) = 1$, $g_1(r_b) = 1$, $g_1(r_c) = 2$, $g_1(r_d) = 3$, $g_1(r_e) = 4$ in the first survey, and $g_2(r_a) = 2$, $g_2(r_b) = 1$, $g_2(r_c) = 3$, $g_2(r_d) = 4$, $g_2(r_e) = 2$ in the second survey. (Note that g_1 and g_2 are specific instances of f_1 and f_2.)

Metric	Intuition	Formal Definition ($\mathcal{R} \coloneqq \mathcal{R}_1 \cap \mathcal{R}_2$)	Running Example on g_1 and g_2
Normalized Undirected Disagreement (NUD)	proportion of resources that moved groups between f_1 and f_2	$\frac{\left\|\left\{r \in \mathcal{R} : f_1(r) \neq f_2(r)\right\}\right\|}{\|\mathcal{R}\|}$	$\mathsf{NUD}(g_1, g_2) = \frac{4}{5} = 0.8$; all resources except r_b moved groups between g_1 and g_2
Normalized Directed Disagreement (NDD)	net proportion of resources that moved forward/backward from f_1 to f_2 (or vice versa)	$\frac{\left\|\left\|\left\{r \in \mathcal{R} : f_1(r) > f_2(r)\right\}\right\| - \left\|\left\{r \in \mathcal{R} : f_1(r) < f_2(r)\right\}\right\|\right\|}{\|\mathcal{R}\|}$	$\mathsf{NDD}(g_1, g_2) = \frac{\|1-3\|}{5} = 0.4$; 1 resource ($r_e$) moved "forward", 3 ($r_a$, r_c, and r_d) moved "backward"
Normalized Pairwise Swap (NPS)	proportion of resource *pairs* whose relative order swapped between f_1 and f_2	$\frac{2 \cdot \left\|\left\{r, r' \in \mathcal{R} : \left(f_1(r) - f_2(r)\right) \cdot \left(f_1(r') - f_2(r')\right) < 0\right\}\right\|}{\|\mathcal{R}\| \cdot (\|\mathcal{R}\| - 1)}$	$\mathsf{NPS}(g_1, g_2) = \frac{2 \cdot 2}{5 \cdot 4} = 0.2$; only two pairs, (r_c, r_e) and (r_d, r_e), had their relative order swapped
Normalized Toggled Resources (NTR)	proportion of resources that are *never used* in one of f_1 and f_2 but not the other	$\frac{\left\|\left\{r \in \mathcal{R} : (f_1(r) = 4, f_2(r) \neq 4) \vee (f_1(r) \neq 4, f_2(r) = 4)\right\}\right\|}{\|\mathcal{R}\|}$	$\mathsf{NTR}(g_1, g_2) = \frac{2}{5} = 0.4$; two resources ($r_d$ and r_e) were "toggled" on/off between g_1 and g_2

two functions is concrete evidence that students' ordering indeed changed regarding the two resources in question, agnostic to any potential context influences.

Finally, there is the question of whether students actually order the available help resources in the granularity that the survey expected. NTR gets to this idea as a cruder version of NUD. More specifically, some students may not order the help resources into multiple sequential groups, and instead only categorize them into *what they use* and *what they do not use* (in other words, merging the first three groups in our surveys into one). NTR is exactly what NUD reduces to in this scenario.

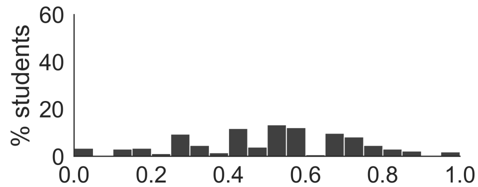
In sum, using all four metrics allows us to more holistically examine whether the same student's two order functions disagree at all (NUD in granularity and NTR without the granularity), and when they do, whether it is due to potential different levels of help needs (NDD) or actual ordering swaps (NPS). In other words, we can more confidently say two functions disagree with each other if they have high NUD and NPS but relatively low NDD and NTR.

Figure 4 illustrates the empirical distribution of each metric mentioned above. Similar to our approach in Section 4.2, for the students that responded to the Order survey in more than two classes (30 out of a total of 260), we obtain the *mean* of each of the four metrics across all pairs.[4] Every data point in Figure 4 is therefore a student, and the figures are the distributions of the means of the metrics. To ensure fair comparison, we only used the mid-semester Order surveys from all classes.

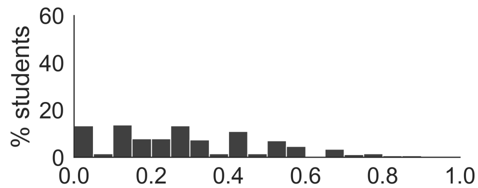
Our results suggest that *what resource(s) a student uses and in what relative order(s)* stay rather consistent for a student in different class contexts. As shown in Figure 4a, the mean NUD values are rather high (both the mean and the median are at 0.50) and are distributed rather evenly between 0 and 1 without obvious spikes, which at the first glance may seem to imply students rank many resources into different groups in different classes. However, students' mean NDD values (plotted in Figure 4b) hints that a substantial part of NUD can be explained by NUD, or in other words, students "shifting the entire distribution forward/backward" according to their different levels of help need. While the distribution of the mean NDD is still meaningfully lower than that of the mean NUD, the mean of students' mean NDD values is 0.30 (which is 60% of that of NUD), suggesting that the disagreement between students' rankings is not as high as what Figure 4a implies. Indeed, as shown in Figure 4c, few pairs of resources (mean = 8%; median = 3%) actually got their relative order swapped between different classes. Note that a student can have nonzero values of NUD and NDD but have a zero value for NPS if they do not completely swap the order of any pair of resources. Finally, Figure 4d shows that relatively few resources (mean = 15%; median = 11%) are used by a student in one class but not in the other, further suggesting that students' *selections of help resources to use* are consistent across courses.

To further contextualize the results in Figure 4, we calculated the same metrics (and plotted their distributions in Figure 5) between each student's response to the mid-semester Order survey and that to the end-of-semester Order survey in all classes that administered both, and found the results to be quite similar to that in Figure 4. These measurements benchmark how much a student's self-reported order of help resource usage could fluctuate in the *same class context* within a timeframe of 1-2 months. As shown in the results, for all four metrics, the empirical distribution of the metric between two responses in the same class (Figure 5) does not substantially differ from that of the mean metric across students' responses in all pairs of their classes (Figure 4). In other words, we observed *students' self-reported order of help resource usage did not meaningfully differ more across courses than within the same courses*, and thus *order of help resource usage is likely a key part of student's individual help-seeking approach.*

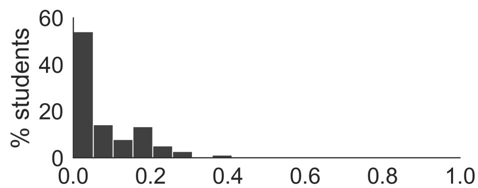
[4] In other words, for a student that filled out the Order survey in classes A, B, and C, their mean NUD would be the mean of three values: their NUD between classes A and B, their NUD between classes A and C, and their NUD between classes B and C.



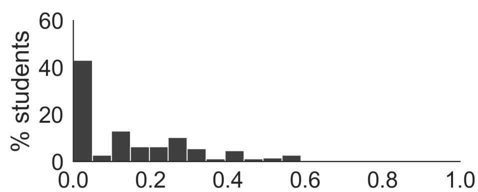
(a) Mean NUD (mean = 0.50; median = 0.50; std = 0.22).



(b) Mean NDD (mean = 0.30; median = 0.25; std = 0.21).

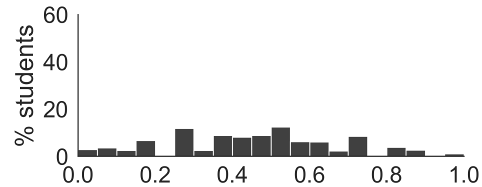


(c) Mean NPS (mean = 0.08; median = 0.03; std = 0.10).

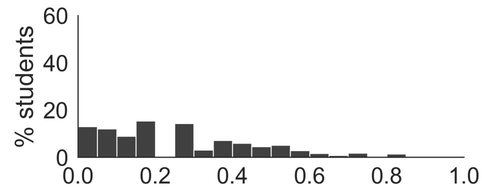


(d) Mean NTR (mean = 0.15; median = 0.11; std = 0.17).

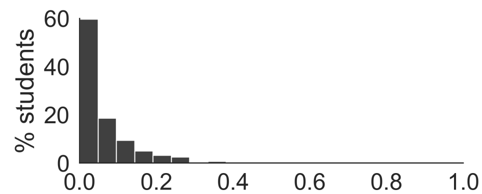
Figure 4: Histograms of mean distance metrics across all pairs of each student's order of help resource usage functions. Data has $N = 260$ students; 230/29/1 students responded to the Order survey in 2/3/4 classes, respectively.



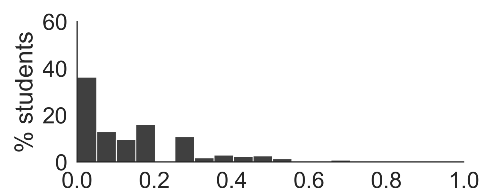
(a) NUD (mean = 0.46; median = 0.45; std = 0.22).



(b) NDD (mean = 0.27; median = 0.27; std = 0.21).



(c) NPS (mean = 0.06; median = 0.04; std = 0.08).



(d) NTR (mean = 0.15; median = 0.14; std = 0.16).

Figure 5: Histograms of distance metrics between students' mid-semester and end-of-semester responses to the Order survey. Data has $N = 929$ entries, each being a student in a class where the Order survey was administered twice.

4.4 Collaboration inclinations *seem to depend on context*

We found students' collaboration inclinations to be more context-driven than individual-driven in our data.

A total of 273 students responded to the Collab survey in more than two courses. For each of these 273 students, we identified their first and last classes taken using the same procedure in Section 4.1. We then ran Spearman's ρ test on their responses to the Collab question across different courses. Note that students' responses to the Likert question (from "*Strongly disagree*" to "*Strongly agree*") are ordinal, enabling us to rank the responses without converting them to numbers. This test did not reveal any significant correlation ($\rho = 0.07, p = 0.26$) between students' collaboration inclinations in their earlier and later classes.

However, different classes have slightly different collaborative cultures in the small-section environments. To control for class-specific context, we further normalized students' responses by order percentiles *per class*, which is again feasible due to the Likert options are ordinal. The Spearman's ρ test on the percentiles showed similar results to that on the raw data ($\rho = 0.11, p = 0.08$).

In sum, we did not find any significant relation between students' responses to the Collab survey across different courses, as the effect sizes were positive but weak (and insignificant) for both analyses. In other words, how collaboratively a student works with others in small-scale discussion/lab sections does not strongly correlate from class to class, and thus should not be treated as part of the student's help-seeking approach.

4.5 Could resource availability have affected order of usage?

By administering the Order survey, we sought to gain more insight on students' help resource utilization patterns in a *different* lens than just looking at students' (self-reported or platform-recorded) frequency of using each formal resource. We thus phrased the survey to specifically ask for students' ideal order of resource usage *assuming all resources are available*, which makes it fundamentally different than existing qualitative works on a similar question [12, 60] in which students' *typical* order was the main focus. In other words, the Order survey attempts to *abstract away* all resource availability contexts, although still using a contextualized list of course-specific help resources. As such, *the responses to the Order survey (students' ideal order of resource usage) should not be directly compared against their order of resource usage in their actual help-seeking records*, because the former is context-agnostic and the latter is contextualized.

However, to explore whether there are significant discrepancies between the two versions of resource usages (which would imply the ecosystems being observed are not ideal for the students), we still attempted a crude comparison in a few classes whose help ecosystems were more conducive to examining the order of resource usage in students' actual records. Specifically, in both CS1-Sp23 and DS-Sp23, our data satisfies the following: (i) the Order survey was administered; (ii) all forum posts and consulting hour interactions were annotated (by students themselves) with the specific course component (granular to the assignment level); (iii) there was clear delineation between UTA-hosted consulting hours vs. GTA/instructor hosted ones (see Table 6 in the Appendix); and (iv) student identities could be linked across different data sources. Therefore, we were able to observe all occurrences in which a student (in either of these two classes) utilized *more than one formal help resources* for the same course component (e.g., a specific assignment), then compare whether their order of utilizing the resources match their self-reported ideal order in the survey.

Our results showed that only 13.5% (50 out of 369) of CS1-Sp23 observations and 16.0% (50 out of 313) of DS-Sp23 observations showed a mismatch between students' Order survey response and their actual usage, i.e., student utilized help resources in a different order than what they indicated in the survey. Note that only when a student accessed two or more formal help resources for the same assignment would it count as an observation.

Furthermore, 40% (20/50) of the CS1-Sp23 discrepancies and 54% (27/50) of the DS-Sp23 discrepancies were occurrences in which student utilized an alternative resource *during a time when the resource that they would prefer to utilize first was not available.* Since the more preferred resource was *not* available when these help utilization happened (i.e., the situation was *not ideal*), these students technically did not deviate from their self-reported *ideal* order of resource usage.

In sum, our preliminary comparison shows that (i) students' actual order of utilizing formal resources mostly match their self-reported ideal order, and (ii) a significant part of discrepancies might be attributed to availability reasons.

5 LIMITATIONS AND FUTURE WORK

Scope of study. Our data on students' help-seeking records only include *formal* (i.e., course-affiliated) platforms and does not capture informal help-seeking such as the frequency students seek help from friends/classmates, online resources, LLMs, etc., which is nonetheless a substantial part of their help-seeking approach. Furthermore, the six courses in this study are all at the same institution and have rather consistent instructional context (e.g., a prominent culture of UTAs, consulting hours are available in the evenings, etc.) that may have contributed to the overall consistent trends in the data. Due to platform and/or logistic inconsistencies, not all kinds of data was collected for every class.

Finally, this study does not capture the fundamental questions of *whether/when students decide to seek help* or *is the help effective when they seek help.* Instead, the study only focuses on *what students do exactly when they seek help*, in that we characterize their help-seeking approach by the frequency (of resource usage), the type (of help), the order (of using resources), and their collaborative inclination.

Analyses. Due to the sparseness of the data, we could not simply analyze the intersection of each pair of classes (which would have made the correlational analyses much more robust). Specifically, the number of overlapping enrolling students is under 50 for all but two pairs of classes. Given that around half of students in each class did not seek any formal help throughout the entirety of the class, analyzing separately by pairs of classes would have rendered all sample sizes under 30. We thus chose to pool all metrics together and perform the (noisier) paired tests between students' records in an earlier and a later course that they took.

We chose to normalize both students' frequency-based help metrics and collaboration inclination using percentiles, which is not a perfect normalization technique and is prone to ignore the relative magnitude of observations. Had the data been less sparse, a more robust strategy would be using a more sophisticated model (e.g., a mixed-effects model) to account for potential differences in help demands amongst different offerings of the same course. Nevertheless, our results for both parts stayed consistent before/after normalization, and thus it is unlikely that the normalization caused any systematic error for the purposes of our analyses.

Data Collection Inaccuracies. Each of our data sources has its own limitation on the granularity of data collected. For consulting hours, existing queueing apps do not well-capture the act of a teaching staff member helping multiple students at a time (due to similar help requests) or bouncing back-and-forth between students (leaving time for students to digest the help and work on their own for healthier learning). Some interactions might be accidentally omitted or recorded with an inaccurate duration due to human error. We thus chose not to analyze the total duration across all consulting hours interactions for a student. Finally, not all kinds of help students sought in consulting hours were captured by the (course-specific) categories, and the process of bucketing the interactions into UPIC phases might lose some nuances.

For the class forum, both the number of days active on the forum and the total number of read actions performed by a student are aggregated over the entire semester and are not thread-specific. We therefore were not able to distinguish between general forum use (e.g., for reading logistic announcements) and help-seeking actions for these two metrics, or account for how different teaching teams used the forum in different ways. The data also did not capture the amount of help students requested in each thread/post, as some students preferred to ask one question per thread while others combined several questions into one thread.

For both the Order and the Collab surveys, the granularity of available options per question is limited: students had to group all help resources into four buckets and estimate their collaboration with peers using a 5-option Likert question. Finally, not all students answered all questions in each survey.

Demographic Variables. One interesting direction is to study the relation between demographic variables and the many help-seeking characteristics analyzed in this study. Due to the underlying student body being not very demographically diverse (see Table 4) and the sparsity of data, we plan to study the effects around single demographic variables as well as intersectionality after accumulating more data.

Interaction between help resources. Our study focuses on *individual students* and the (cor-)relation between individual students' help-seeking records/approaches in different contexts. As such, we do not analyze the interaction between *help resources* (e.g., whether the frequencies of getting help on the class forum and in consulting hours are correlated, and in which direction, etc.). We leave such research questions to a future paper on its own.

6 CONCLUSIONS

Our work is a preliminary attempt in characterizing computing students' *individual* help-seeking approaches. Our analyses revealed that frequency-based help metrics and students' ideal order of resource usage stay consistent for individual students across different course contexts, and thus may be treated as part of their individual help-seeking approaches, while type of help sought in consulting hours and whether students collaborate with peers in lab/discussions do not exhibit as much consistency. Although many interesting questions remain, our work adds evidence that individual help-seeking approaches do exist and should be considered an essential part of any model of the help-seeking process while also showing that not every bit of students' help-seeking characteristics is part of their approach.

ACKNOWLEDGMENTS

We thank all instructors and teaching associates of all classes in our dataset for their help in collecting and transferring the data. We thank Jonathan Liu and the anonymous reviewers for their extremely helpful and constructive advices on the presentation of the paper. This work is supported by the National Science Foundation grant #1934965.

REFERENCES

[1] Elisa L. A. Baniassad, Lucas Zamprogno, Braxton Hall, and Reid Holmes. 2021. STOP THE (AUTOGRADER) INSANITY: Regression Penalties to Deter Autograder Overreliance. In *ACM SIGCSE TS*. 1062–1068. https://doi.org/10.1145/3408877.3432430

[2] Soumya Basu, Albert Wu, Brian Hou, and John DeNero. 2015. Problems Before Solutions: Automated Problem Clarification at Scale. In *ACM L@S*. 205–213. https://doi.org/10.1145/2724660.2724679

[3] Brett A. Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. 2023. Programming Is Hard - Or at Least It Used to Be: Educational Opportunities and Challenges of AI Code Generation. In *ACM SIGCSE TS*. 500–506. https://doi.org/10.1145/3545945.3569759

[4] Kathryn Bridson, Jeffrey Atkinson, and Scott D. Fleming. 2022. Delivering Round-the-Clock Help to Software Engineering Students Using Discord: An Experience Report. In *ACM SIGCSE TS*. 759–765. https://doi.org/10.1145/3478431.3499385

[5] Jennifer Campbell and Michelle Craig. 2018. Drop-In Help Centres: An Alternative to Office Hours. In *ACM Western Canadian Conference on Computing Education*. 9:1–9:6. https://doi.org/10.1145/3209635.3209642

[6] Sung-Hyuk Cha. 2007. Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences* 4 (2007), 300–307. https://api.semanticscholar.org/CorpusID:15506682

[7] Chih-Yueh Chou and Chun-Ho Chang. 2021. Developing Adaptive Help-Seeking Regulation Mechanisms for Different Help-Seeking Tendencies. *Educational Technology & Society* 24 (2021), 54–66. https://www.jstor.org/stable/48629244

[8] Elizabeth B. Cloude, Ryan Shaun Baker, and Eric Fouh. 2023. Online help-seeking occurring in multiple computer-mediated conversations affects grades in an introductory programming course. In *ACM LAK*. 378–387. https://doi.org/10.1145/3576050.3576106

[9] Galina Deeva, Daria Bogdanova, Estefanía Serral, Monique Snoeck, and Jochen De Weerdt. 2021. A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Comput. Educ.* 162 (2021), 104094. https://doi.org/10.1016/j.compedu.2020.104094

[10] Andrew Deorio and Christina Keefer. 2022. When is Automated Feedback a Barrier to Timely Feedback?. In *2022 ASEE Annual Conference & Exposition*. https://peer.asee.org/40405

[11] Ed Discussion. 2024. https://edstem.org/.

[12] Augie Doebling and Ayaan M. Kazerouni. 2021. Patterns of Academic Help-Seeking in Undergraduate Computing Students. In *ACM Koli Calling*. 13:1–13:10. https://doi.org/10.1145/3488042.3488052

[13] Pierpaolo Dondio and Suha Shaheen. 2019. Is StackOverflow an Effective Complement to Gaining Practical Knowledge Compared to Traditional Computer Science Learning?. In *ACM ICETC*. 132–138. https://doi.org/10.1145/3369255.3369258

[14] Yang-Hsin Fan and Tzung-Jin Lin. 2023. Identifying university students' online academic help-seeking patterns and their role in Internet self-efficacy. *The Internet and Higher Education* 56 (2023), 100893. https://doi.org/10.1016/j.iheduc.2022.100893

[15] Sara J. Finney, Carol L. Barry, S. Jeanne Horst, and Mary M. Johnston. 2018. Exploring profiles of academic help seeking: A mixture modeling approach. *Learning and Individual Differences* 61 (2018), 158–171. https://doi.org/10.1016/j.lindif.2017.11.011

[16] Carlton J. Fong, Cassandra Gonzales, Christie Hill-Troglin Cox, and Holly B. Shinn. 2023. Academic help-seeking and achievement of postsecondary students: A meta-analytic investigation. *Journal of Educational Psychology* 115 (2023), 1–21. https://doi.org/10.1037/edu0000725

[17] 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*. 508–514. https://doi.org/10.5281/zenodo.6853172

[18] Zhikai Gao, Sarah Heckman, and Collin Lynch. 2022. Who Uses Office Hours?: A Comparison of In-Person and Virtual Office Hours Utilization. In *ACM SIGCSE TS*. 300–306. https://doi.org/10.1145/3478431.3499334

[19] Zhikai Gao, Collin Lynch, Sarah Heckman, and Tiffany Barnes. 2021. Automatically Classifying Student Help Requests: A Multi-Year Analysis. In *EDM*. 81–92. https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_153.pdf

[20] Joan Giblin, Jill E. Stefaniak, Angela Eckhoff, and Tian Luo. 2021. An exploration of factors influencing the decision-making process and selection of academic help sources. *J. Comput. High. Educ.* 33, 1 (2021), 1–18. https://doi.org/10.1007/s12528-020-09252-0

[21] Georgiana Haldeman, Monica Babes-Vroman, Andrew Tjang, and Thu D. Nguyen. 2021. CSF: Formative Feedback in Autograding. *ACM Trans. Comput. Educ.* 21, 3 (2021), 21:1–21:30. https://doi.org/10.1145/3445983

[22] Qiang Hao, David H. Smith IV, Lu Ding, Amy J. Ko, Camille Ottaway, Jack Wilson, Kai Arakawa, Alistair Turcan, Timothy Poehlman, and Tyler Greer. 2022. Towards understanding the effective design of automated formative feedback for programming assignments. *Comput. Sci. Educ.* 32, 1 (2022), 105–127. https://doi.org/10.1080/08993408.2020.1860408

[23] Arto Hellas, Juho Leinonen, Sami Sarsa, Charles Koutcheme, Lilja Kujanpää, and Juha Sorva. 2023. Exploring the Responses of Large Language Models to Beginner Programmers' Help Requests. In *ACM ICER*. 93–105. https://doi.org/10.1145/3568813.3600139

[24] Andrew D. Hilton, Genevieve M. Lipp, and Susan H. Rodger. 2019. Translation from Problem to Code in Seven Steps. In *ACM CompEd*. 78–84. https://doi.org/10.1145/3300115.3309508

[25] Irene Hou, Sophia Mettille, Owen Man, Zhuo Li, Cynthia Zastudil, and Stephen MacNeil. 2024. The Effects of Generative AI on Computing Students' Help-Seeking Preferences. In *ACM ACE*. 39–48. https://doi.org/10.1145/3636243.3636248

[26] Stuart A. Karabenick. 2003. Seeking help in large college classes: A person-centered approach. *Contemporary educational psychology* 28, 1 (2003), 37–58. https://doi.org/10.1016/S0361-476X(02)00012-7

[27] Stuart A. Karabenick. 2004. Perceived achievement goal structure and college student help seeking. *Journal of educational psychology* 96, 3 (2004), 569. https://doi.org/10.1037/0022-0663.96.3.569

[28] Stuart A. Karabenick and Jean-Louis Berger. 2013. Help seeking as a self-regulated learning strategy. In *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman*, Héfer Bembenutty, Timothy J. Cleary, and Anastasia Kitsantas (Eds.). 237–261. https://www.infoagepub.com/products/Applications-of-Self-Regulated-Learning-across-Diverse-Disciplines

[29] 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

[30] Stuart A. Karabenick and John R. Knapp. 1991. Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of educational psychology* 83, 2 (1991), 221. https://doi.org/10.1037/0022-0663.83.2.221

[31] Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J. Ericson, David Weintrop, and Tovi Grossman. 2023. Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In *ACM CHI*. 455:1–455:23. https://doi.org/10.1145/3544548.3580919

[32] Hieke Keuning, Johan Jeuring, and Bastiaan Heeren. 2019. A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. *ACM Trans. Comput. Educ.* 19, 1 (2019), 3:1–3:43. https://doi.org/10.1145/3231711

[33] Shao-Heng Ko and Kristin Stephens-Martinez. 2023. What drives students to office hours: individual differences and similarities. In *ACM SIGCSE TS*. 959–965. https://doi.org/10.1145/3545945.3569777

[34] Sophia Krause-Levy, Rachel S. Lim, Ismael Villegas Molina, Yingjun Cao, and Leo Porter. 2022. An Exploration of Student-Tutor Interactions in Computing. In *ACM ITiCSE*. 435–441. https://doi.org/10.1145/3502718.3524786

[35] Harrison Kwik, Haoqi Zhang, and Eleanor O'Rourke. 2022. How Do Students Seek Help and How Do TAs Respond?: Investigating Help-Seeking Strategies in CS1 Office Hours. In *ACM SIGCSE TS*. 1130. https://doi.org/10.1145/3478432.3499099

[36] Sam Lau and Philip J. Guo. 2023. From "Ban It Till We Understand It" to "Resistance is Futile": How University Programming Instructors Plan to Adapt as More Students Use AI Code Generation and Explanation Tools such as ChatGPT and GitHub Copilot. In *ACM ICER*. 106–121. https://doi.org/10.1145/3568813.3600138

[37] Juho Leinonen, Paul Denny, Stephen MacNeil, Sami Sarsa, Seth Bernstein, Joanne Kim, Andrew Tran, and Arto Hellas. 2023. Comparing Code Explanations Created by Students and Large Language Models. In *ACM ITiCSE*. 124–130. https://doi.org/10.1145/3587102.3588785

[38] Juho Leinonen, Paul Denny, and Jacqueline Whalley. 2022. A Comparison of Immediate and Scheduled Feedback in Introductory Programming Projects. In *ACM SIGCSE TS*. 885–891. https://doi.org/10.1145/3478431.3499372

[39] Rongxin Liu, Carter Zenke, Charlie Liu, Andrew Holmes, Patrick Thornton, and David J. Malan. 2024. Teaching CS50 with AI: Leveraging Generative Artificial Intelligence in Computer Science Educatgion. In *ACM SICCSE TS*. 750–756. https://doi.org/10.1145/3626252.3630938

[40] Dastyni Loksa, Lauren E. Margulieux, Brett A. Becker, Michelle Craig, Paul Denny, Raymond Pettit, and James Prather. 2022. Metacognition and Self-Regulation in Programming Education: Theories and Exemplars of Use. *ACM Trans. Comput. Educ.* 22, 4 (2022), 39:1–39:31. https://doi.org/10.1145/3487050

[41] Stephen MacNeil, Andrew Tran, Juho Leinonen, Paul Denny, Joanne Kim, Arto Hellas, Seth Bernstein, and Sami Sarsa. 2023. Automatically Generating CS Learning Materials with Large Language Models. In *ACM SIGCSE TS*. 1176. https://doi.org/10.1145/3545947.3569630

[42] Kara A. Makara and Stuart A. Karabenick. 2013. Characterizing sources of academic help in the age of expanding educational technology: A new conceptual framework. In *Advances in help-seeking research and applications: The role of emerging technologies*, Stuart A. Karabenick and Minna Puustinen (Eds.). 37–72. https://www.infoagepub.com/products/Advances-in-Help-Seeking-Research-and-Applications

[43] Samiha Marwan, Ge Gao, Susan R. Fisk, Thomas W. Price, and Tiffany Barnes. 2020. Adaptive Immediate Feedback Can Improve Novice Programming Engagement and Intention to Persist in Computer Science. In *ACM ICER*. 194–203. https://doi.org/10.1145/3372782.3406264

[44] MyDigitalHand. 2024. https://beta.mydigitalhand.org/.

[45] Piazza. 2024. https://piazza.com/.

[46] James Prather, Lauren Margulieux, Jacqueline Whalley, Paul Denny, Brent N. Reeves, Brett A. Becker, Paramvir Singh, Garrett Powell, and Nigel Bosch. 2022. Getting By With Help From My Friends: Group Study in Introductory Programming Understood as Socially Shared Regulation. In *ACM ICER*. 164–176. https://doi.org/10.1145/3501385.3543970

[47] James Prather, Brent N. Reeves, Paul Denny, Brett A. Becker, Juho Leinonen, Andrew Luxton-Reilly, Garrett Powell, James Finnie-Ansley, and Eddie Antonio Santos. 2023. "It's Weird That It Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. *ACM Trans. Comput.-Hum. Interact.* (2023). https://doi.org/10.1145/3617367

[48] Thomas W. Price, Zhongxiu Liu, Veronica Cateté, and Tiffany Barnes. 2017. Factors Influencing Students' Help-Seeking Behavior while Programming with Human and Computer Tutors. In *ACM ICER*. 127–135. https://doi.org/10.1145/3105726.3106179

[49] Adnan Qayyum. 2018. Student help-seeking attitudes and behaviors in a digital era. *International Journal of Educational Technology in Higher Education* 15, 1 (2018), 1–16. https://doi.org/10.1186/s41239-018-0100-7

[50] Yanyan Ren, Shriram Krishnamurthi, and Kathi Fisler. 2019. What Help Do Students Seek in TA Office Hours?. In *ACM ICER*. 41–49. https://doi.org/10.1145/3291279.3339418

[51] Christian Schlusche, Lenka Schnaubert, and Daniel Bodemer. 2023. Understanding students' academic help-seeking on digital devices - a qualitative analysis. *Research and Practice in Technology Enhanced Learning* 18 (2023), 17. https://doi.org/10.58459/rptel.2023.18017

[52] Mrinal Sharma, Hayden McTavish, Zimo Peng, Anshul Shah, Vardhan Agarwal, Caroline Sih, Emma Hogan, Ismael Villegas Molina, Adalbert Gerald Soosai Raj, and Kristen Vaccaro. 2023. Engagement and Anonymity in Online Computer Science Course Forums. In *ACM ICER*. 48–62. https://doi.org/10.1145/3568813.3600121

[53] 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 *ACM SIGCSE TS*. 549–554. https://doi.org/10.1145/3017680.3017800

[54] David H Smith IV, Qiang Hao, Vanessa Dennen, Michail Tsikerdekis, Bradly Barnes, Lilu Martin, and Nathan Tresham. 2020. Towards Understanding Online Question & Answer Interactions and their effects on student performance in large-scale STEM classes. *International Journal of Educational Technology in Higher Education* 17 (2020), 1–15. https://doi.org/10.1186/s41239-020-00200-7

[55] Nick A Stites, Edward Berger, Jennifer Deboer, and Jeffrey F Rhoads. 2019. A cluster-based approach to understanding students' resource-usage patterns in an active, blended, and collaborative learning environment. *International Journal of Engineering Education* 35, 6 (2019), 1738–1757. https://www.ijee.ie/1atestissues/Vol35-6A/16_ijee3834.pdf

[56] Sadhana Suryadevara and Kristin Stephens-Martinez. 2022. UPIC a Problem-Solving Framework: Understand, Plan, Implement, and Correctness/Debugging. In *ACM ICER*. 50–51. https://doi.org/10.1145/3501709.3544286

[57] Adrian Thinnyun, Ryan Lenfant, Raymond Pettit, and John R. Hott. 2021. Gender and Engagement in CS Courses on Piazza. In *ACM SIGCSE TS*. 438–444. https://doi.org/10.1145/3408877.3432395

[58] Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik. 2014. Gender, representation and online participation: A quantitative study. *Interacting with Computers* 26, 5 (2014), 488–511. https://doi.org/10.1093/iwc/iwt047

[59] Mickey Vellukunnel, Philip Buffum, Kristy Elizabeth Boyer, Jeffrey Forbes, Sarah Heckman, and Ketan Mayer-Patel. 2017. Deconstructing the Discussion Forum: Student Questions and Computer Science Learning. In *ACM SIGCSE TS*. 603–608. http://doi.org/10.1145/3017680.3017745

[60] Elizabeth Wirtz, Amy Dunford, Edward Berger, Elizabeth Briody, Gireesh Guruprasad, and Ryan Senkpeil. 2018. Resource usage and usefulness: academic help-seeking behaviours of undergraduate engineering students. *Australasian Journal of Engineering Education* 23, 2 (2018), 62–70. http://doi.org/10.1080/22054952.2018.1525889

[61] Jack Wrenn and Shriram Krishnamurthi. 2021. Reading Between the Lines: Student Help-Seeking for (Un)Specified Behaviors. In *ACM Koli Calling*. 14:1–14:6. http://doi.org/10.1145/3488042.3488072

[62] 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

[63] Matthew Zahn and Sarah Heckman. 2023. Observations on Student Help-Seeking Behaviors in Introductory Computer Science Courses. In *ACM SIGCSE TS*. 1380. http://doi.org/10.1145/3545947.3576325

[64] Barry J. Zimmerman. 2001. Theories of self-regulated learning and academic achievement: An overview and analysis. In *Self-regulated learning and academic achievement: Theoretical perspectives*, Barry J. Zimmerman and Dale H. Schunk (Eds.). 1–37. https://doi.org/10.1007/978-1-4612-3618-4

[65] Zoom. 2024. https://zoom.us/.

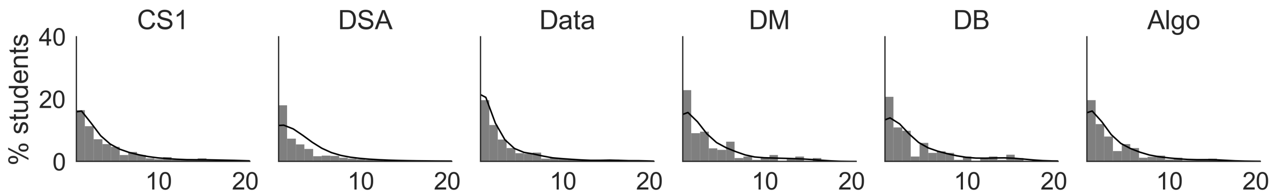
A SUPPLEMENTARY MATERIAL FOR SECTION 3

Table 4: Student demographics (consenting students only). Not all consenting students responded to the demographics survey; in classes where the demographics were collected in a separate survey, the number of consenting responses to the survey could be less than the number of consenting students. Black includes African American, and 2+ stands for multiracial. Values lower than 5 are replaced with an asterisk. Not all data was collected in all classes, and not all students answered all questions.

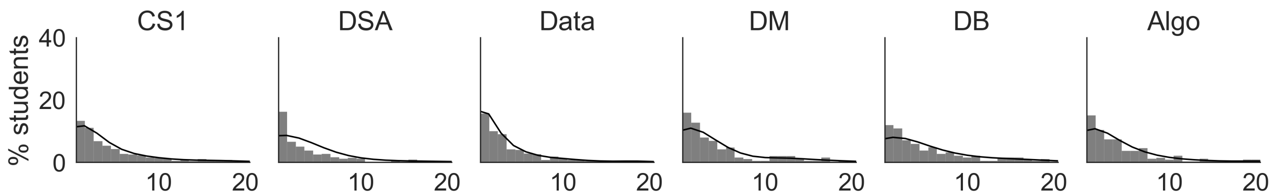
Course-	Semester	# consenting		Gender			Race					Hispanic	Year			
		students	(responses)	Men	Women	Nonbinary	White	Asian	Black	2+	Other		1	2	3	4
CS1	Fa20	152	152	75	73	*	74	42	12	13	*	19	-	-	-	-
	Sp21	157	141	47	93	*	65	50	8	11	*	15	92	22	22	5
	Fa21	177	161	64	94	*	80	48	15	13	*	22	87	48	14	11
	Sp22	152	124	48	76	*	63	38	*	16	*	22	81	27	10	6
	Fa22	183	175	-	-	-	-	-	-	-	*	-	106	32	23	14
	Sp23	163	163	63	90	*	90	35	11	15	*	20	101	25	16	16
DSA	Sp21	217	212	136	75	*	94	81	13	14	*	22	143	53	12	*
	Sp23	238	238	125	110	*	89	97	20	15	*	31	176	39	18	*
	Fa23	272	272	167	97	5	100	131	14	13	14	24	157	93	15	5
Data	Sp21	181	181	79	93	*	76	69	12	12	*	16	10	40	68	57
	Fa21	144	144	78	62	*	61	55	9	10	*	10	*	30	60	48
	Sp22	145	145	84	55	*	59	63	8	5	*	11	10	56	46	27
	Fa22	152	152	83	65	*	41	83	8	12	*	14	*	34	54	54
	Sp23	160	160	92	61	*	63	67	7	15	*	23	21	68	50	20
	Fa23	66	66	33	32	*	22	33	*	6	*	7	*	19	18	29
DM	Sp23	94	62	33	27	*	18	29	6	5	*	8	14	25	22	*
	Fa23	94	94	49	42	*	23	44	5	13	7	7	5	42	37	10
DB	Fa23	183	183	104	74	*	61	97	6	9	*	21	*	34	75	68
Algo	Sp23	202	202	120	77	*	89	88	12	11	*	18	*	44	91	67
	Fa23	119	119	83	34	*	45	59	8	5	*	11	*	9	36	65

Table 5: Mapping from class-specific problem-solving step/phases to UPIC [56] (Understand, Plan, Implement, and Correctness) phases. Note that only CS1, DSA, and Data collected the type of help sought in consulting hours.

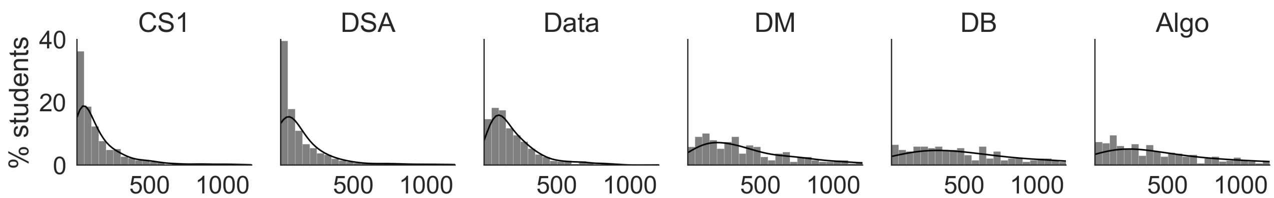
Course	Option	Understand	Plan	Implement	Correctness
CS1	Doing an instance of the problem (Step 1 of the 7-steps [24])	×			
	Developing a plan to solve a problem (Steps 3 and 4 of the 7-steps)		×		
	Writing the code to solve a problem (Step 5 of the 7-steps)			×	
	Testing my program (Step 6 of the 7-steps)				×
DSA	Understanding a problem or directions	×			
	Reviewing past materials/concepts	×			
	Planning how to solve a problem / thinking about what code to write		×		
	Implementing my plan / writing my code			×	
	Testing / understanding my code and debugging				×
Data	Understanding a problem or directions	×			
	Understanding a concept from class	×			
	Planning how to solve a problem before getting into the math/code details		×		
	Writing the math/code details to solve a problem			×	
	Validating/testing/debugging my solution				×



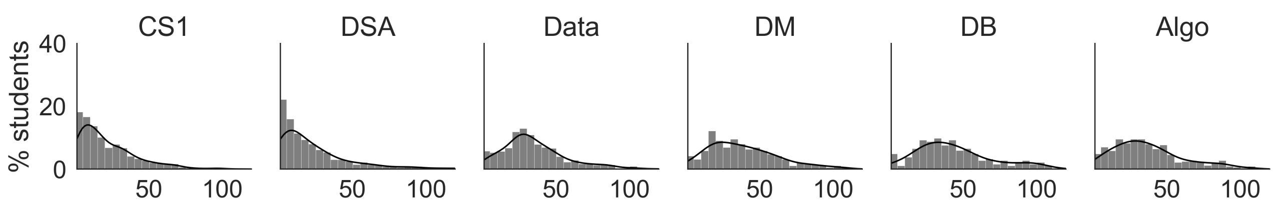
(a) Number of Class Forum Threads Initiated.



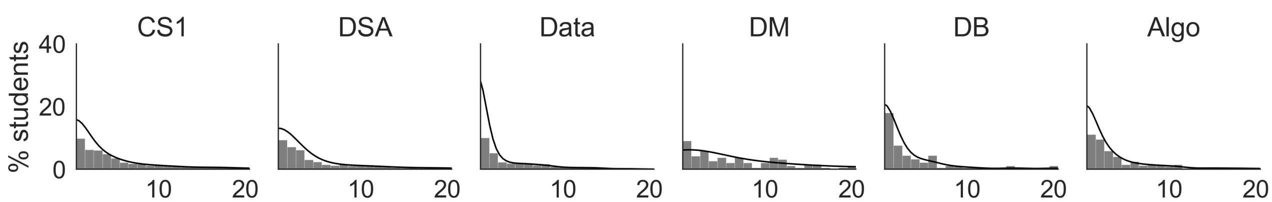
(b) Number of Class Forum Contributions (Threads/Comments/Answers).



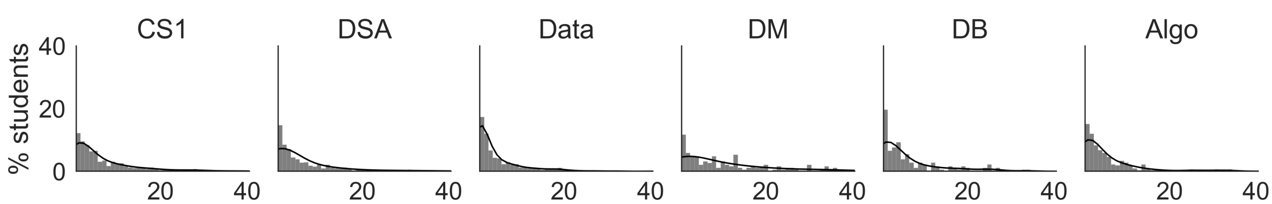
(c) Number of Total Views on Class Forums.



(d) Number of Days Active on Class Forums.



(e) Number of Consulting Hours Interactions.

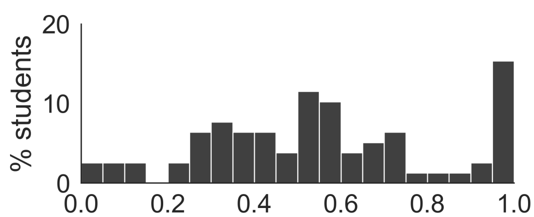


(f) Number of Social Help Requests (sum of Parts 6a and 6e).

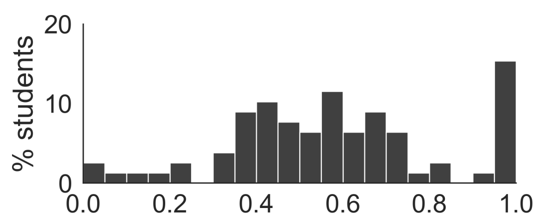
Figure 6: Histograms and KDEs for frequency-based help metrics for each course.

Table 6: Help resources included in each class's Order survey. Note that not all classes had the Order survey (See Table 1). An × in a cell indicates the help resource was included in the Order survey as a standalone option. An - indicates the help resource was not part of the class's help ecosystem. Otherwise, resources with the same symbols were captured as the same option in the survey. For example, in DSA-Sp23, Data-Sp23, Data-Fa23, and DB-Fa23, there was an option in the Order survey phrased as "GTA/Instructor online consulting hours" that captured both resources with the △ symbol.

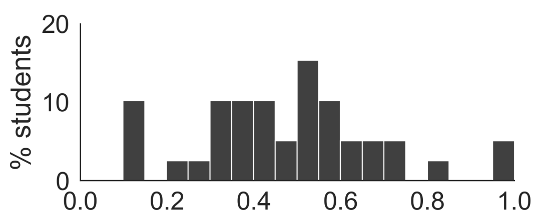
Class	CS1	DSA		Data			DM		DB	Algo	
Resources	Sp23	Sp23	Fa23	Fa22	Sp23	Fa23	Sp23	Fa23	Fa23	Sp23	Fa23
Class material	×	×	×	×	×	×	×	×	×	×	×
Online resources-static	†	†	×	†	†	×	†	×	×	†	×
Online resources-LLMs	†	†	×	†	†	×	†	×	×	†	×
Classmates	×	×	×	×	×	×	×	×	×	×	×
Class forum-read	×	×	×	‡	×	×	×	×	×	×	×
Class forum-write	×	×	×	‡	×	×	×	×	×	×	×
UTA in-person consulting hours	×	×	×	◄	×	×	◊	►	×	►	►
UTA online consulting hours	×	×	×	◄	×	×	◊	▷	-	▷	▷
GTA in-person consulting hours	▲	▲	-	◁	▲	-	◊	►	-	►	►
GTA online consulting hours	-	△	-	◁	△	△	◊	▷	△	▷	▷
Instructor in-person consulting hours	▲	▲	×	♦	▲	×	♦	×	-	×	×
Instructor online consulting hours	×	△	×	♦	△	△	♦	-	△	×	×
People unaffiliated	×	×	×	×	×	×	×	×	×	×	×



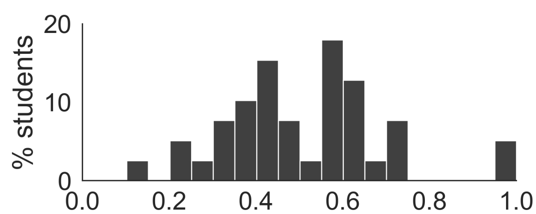
(a) $\overline{d}_{\mathrm{CB}}(stu)$, all students ($N = 78$).



(b) $\overline{d}_{\mathrm{H}}(stu)$, all students ($N = 78$).



(c) $\overline{d}_{\mathrm{CB}}(stu)$, heavy users ($N = 39$).



(d) $\overline{d}_{\mathrm{H}}(stu)$, heavy users ($N = 39$).

Figure 7: Distributions of mean distance metrics between all pairs of each student's UPIC distributions. Data has $N = 78$ students; 68/9/1 students had at least one consulting hours interaction captured by the UPIC framework in 2/3/4 classes, respectively. 39 out of the 78 students had at least 3 UPIC-captured interactions in multiple classes.

B SUPPLEMENTARY MATERIAL FOR SECTION 4.2

We present below the formal definitions of mean Cityblock and Hellinger distances used in the analyses in Section 4.2, then present the results for both to allow for comparison.

Define the normalized empirical UPIC vector of each student stu's consulting hours usage history in each class cls as

$$\mathbf{P}_{\mathsf{UPIC}}(stu, cls) := \langle p_{\mathsf{u}}, p_{\mathsf{p}}, p_{\mathsf{i}}, p_{\mathsf{c}} \rangle,$$

where

$$p_{\mathsf{x}} := \frac{\text{\# of } stu\text{'s interactions in } cls \text{ in phase x}}{\text{\# of all } stu\text{'s interactions in } cls \text{ captured by the UPIC framework}},$$

for all $\mathsf{x} \in \{\mathsf{u}, \mathsf{p}, \mathsf{i}, \mathsf{c}\}$.

Note that the numerator of p_{x} is not always an integer in some classes: among the classes in which the type of help question was select-all-that-apply, 11.4%-25.1% of the interactions were annotated with multiple UPIC phases. For those interactions, we distributed an equal weight to each selected phase. For example, an interaction for which the student selected Understand, Plan, and Implement will count as 1/3 of an interaction to each of the three phases.

The Cityblock and Hellinger distances [6] between two empirical UPIC vectors $\mathbf{P} = \langle p_{\mathsf{u}}, p_{\mathsf{p}}, p_{\mathsf{i}}, p_{\mathsf{c}} \rangle$ and $\mathbf{Q} = \langle q_{\mathsf{u}}, q_{\mathsf{p}}, q_{\mathsf{i}}, q_{\mathsf{c}} \rangle$ are defined as

$$d_{\mathsf{CB}}(\mathbf{P}, \mathbf{Q}) := \frac{1}{2} \left(\sum_{\mathsf{x} \in \{\mathsf{u,p,i,c}\}} |p_{\mathsf{x}} - q_{\mathsf{x}}| \right),$$

$$d_{\mathsf{H}}(\mathbf{P}, \mathbf{Q}) := \sqrt{1 - \left(\sum_{\mathsf{x} \in \{\mathsf{u,p,i,c}\}} \sqrt{p_{\mathsf{x}} \cdot q_{\mathsf{x}}} \right)}.$$

Both distance metrics have a range of 0 to 1, where a distance of 0 happens when the two empirical UPIC vectors are identical, and a distance of 1 happens when the two empirical distributions are completely disjoint.

The two metrics measure discrepancies between empirical probability vectors differently. For a concrete example, suppose $\mathbf{P} = \langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \rangle$ (i.e., a uniform distribution) and $\mathbf{Q} = \langle \frac{1}{2}, \frac{1}{2}, 0, 0 \rangle$ (i.e., a uniform distribution across only the first two phases). Accordingly, the Cityblock distance puts equal weight on the discrepancies between each phase's weights in the two distributions (all four terms are $\frac{1}{4}$). On the other hand, the Hellinger distance deems the two distributions *more discrepant* in the later two phases than the first two: since the Implement phase has a positive weight in $\mathbf{P}$ but no weight in $\mathbf{Q}$, the two distributions are drastically different regarding this phase, as opposed to that the Understand phase contributing a weight of $\sqrt{\frac{1}{4} \cdot \frac{1}{2}}$ to the negative term.

Each student's mean Cityblock and Hellinger distances are then defined as

$$\overline{d}_{\mathsf{CB}}(stu) := \frac{\sum\limits_{c,c' \in C(stu),\, c \neq c'} d_{\mathsf{CB}}\Big(\mathbf{P}_{\mathsf{UPIC}}(stu, c), \mathbf{P}_{\mathsf{UPIC}}(stu, c')\Big)}{|C(stu)| \cdot (|C(stu)| - 1)},$$

$$\overline{d}_{\mathsf{H}}(stu) := \frac{\sum\limits_{c,c' \in C(stu),\, c \neq c'} d_{\mathsf{H}}\Big(\mathbf{P}_{\mathsf{UPIC}}(stu, c), \mathbf{P}_{\mathsf{UPIC}}(stu, c')\Big)}{|C(stu)| \cdot (|C(stu)| - 1)},$$

where $C(stu)$ is the set of classes that student stu participated in. As such, both mean distances $\overline{d}_{\mathsf{CB}}(stu)$ and $\overline{d}_{\mathsf{H}}(stu)$ are still between 0 and 1 for each student stu, capturing (in aggregate) how similar/different each student's consulting hours usage records across all of their classes are, in terms of problem-solving phases.

Figure 7 shows the distribution of both metrics. The results using Hellinger distances are similar to that of the Cityblock distances: there is a peak at 1.0 (students with sparse records), and the rest of the values are still high (median = 0.56 for all students and median = 0.48 for heavy users). In other words, regardless of which metric we use, our results reveal a substantial distributional difference across students' own consulting hours usage records in terms of their problem-solving phases, which seems to be driven by instructional context.