Rethinking Computing Students' Help Resource Utilization through Sequentiality

SHAO-HENG KO and KRISTIN STEPHENS-MARTINEZ, Department of Computer Science, Duke University, Durham, North Carolina, USA

Background. Academic help-seeking benefits students' achievement, but existing literature either studies important factors in students' selection of all help resources via self-reported surveys or studies their help-seeking behavior in one or two separate help resources via actual help-seeking records. Little is known about whether computing students' approaches and behavior match, and not much is understood about how they transition sequentially from one help resource to another.

Objectives. We aim to study post-secondary computing students' academic help-seeking approach and behavior. Specifically, we seek to investigate students' self-reported orders of resource usage and whether these approaches match with students' actual utilization of help resources. We also examine frequent patterns emerging from students' chronological help-seeking records in course-affiliated help resources.

Context and Study Method. We surveyed students' self-reported orders of resource usage across 12 offerings of seven courses at two institutions, then analyzed their responses using various help resource dimensions identified by existing works. From two of these courses (an introduction to programming course and a data science course, 11 offerings), we obtained students' help-seeking records in all course-affiliated help resources, along with code autograder records. We then compared students' reported orders in these two courses against their actions in the records. Finally, we mined sequences of student help-seeking events from these two courses to reveal frequent sequential patterns.

Findings. Students' reported orders of help resource usage form a progression of clusters where resources in each cluster are more similar to each other by help resource dimensions than to resources outside of their cluster. This progression partially confirms phenomena and decision factors reported by existing literature, but no factor/dimension alone can explain the entire progression. We found students' actual help-seeking records did not deviate much from their self-reported orders. Mining of the sequential records revealed that help-seeking from course-affiliated human resources led to measurable progress more often than not, and students' usage of consulting/office hours (mainly run by undergraduate teaching assistants) itself was the best indicator for future usage within the lifespan of the same assignment.

Implications. Our results demonstrate that computing students' help resource selection/utilization is a sophisticated process that should be modeled and analyzed with sufficient awareness of its inherent sequentiality. We identify future research directions through this preliminary analysis, which can lead to a better understanding of computing students' help-seeking behavior and better resource utilization/management in large-scale instructional contexts.

CCS Concepts: • **Social and professional topics → Computing education**;

This work is supported by the National Science Foundation, Division of Computer and Network Systems Grants #1934965 and #1934975, and Division of Undergraduate Education Grants #2336804 and #2336805.

Authors' Contact Information: Shao-Heng Ko (corresponding author), Department of Computer Science, Duke University, Durham, North Carolina, USA; e-mail: shaoheng.ko@duke.edu; Kristin Stephens-Martinez, Department of Computer Science, Duke University, Durham, North Carolina, USA; e-mail: ksm@cs.duke.edu.

This work is licensed under a Creative Commons Attribution-ShareAlike International 4.0 License.

© 2025 Copyright held by the owner/author(s).

ACM 1946-6226/2025/4-ART7
https://doi.org/10.1145/3716860

Additional Key Words and Phrases: Computing Education, Help-Seeking

ACM Reference format:
Shao-Heng Ko and Kristin Stephens-Martinez. 2025. Rethinking Computing Students' Help Resource Utilization through Sequentiality. *ACM Trans. Comput. Educ.* 25, 1, Article 7 (April 2025), 34 pages.
https://doi.org/10.1145/3716860

1 Introduction

Academic help-seeking is a metacognitive and **self-regulated learning (SRL)** strategy employed by students [38, 53, 82]. Instrumental help-seeking, which emphasizes acquiring mastery rather than results, benefits students' academic achievement [17]. A key stage in students' *help-seeking process* [39] is that of *resource selection* [26]: After identifying the need for help but before soliciting help, students must first *decide on* the help resource(s) to solicit help from. Starting from Makara and Karabenick's seminal work [55], this decision-making process has been studied both in the general post-secondary context [26, 70, 73] and in engineering/computing education contexts [14, 33, 69, 82, 89]. These works revealed many factors that influence students' resource selection including *accessibility/availability of the resource* [14, 26, 55, 69, 73, 89], *perceived usefulness of the resource* [14, 55, 89], *formality of the resource* [14, 89], *synchronicity/timeliness of response* [26, 73], and *required effort to phrase a problem* [73].

The above factors, when viewed as different *dimensions*, collectively form a rich *help resource landscape* [46]: Each help resource is a point in the landscape, and students can proactively *transition* from point to point to effectively seek desired help. For a concrete example, **consulting hours (CH)**, also called *office hours* or *tutoring hours* in past works, is a designated time/place where students get helped by an instructor or a **teaching assistant (TA)** on a *one-on-one* basis. We may categorize an in-person CH run by an instructor as *formal*, *highly synchronous*, and *semi-accessible* (due to being limited to a certain time and location). On the other hand, an online class forum/discussion board would be *semi-formal*, *mostly asynchronous*, and *very accessible*.

However, the help ecosystems around modern computing classrooms have been evolving at a higher speed than what existing literature has kept up with. The gaps in existing theories are threefold:

New Resources and Dimensions in the Help Landscape. Due to the growth of enrollment in computing-related fields, an increasing number of large-scale computing classes now heavily rely on **undergraduate teaching assistants (UTAs)** as a help resource [57, 59]. Such help is not yet well-captured by the *formality* dimension previously used as a dichotomy [17, 55], as UTAs can be seen as both formal (affiliated with the course) and informal (as similar-aged peers).

Meanwhile, *code autograders* (*autograders* here on) are now heavily used as a (formal, highly accessible, and synchronous) field-specific help resource to give constructive and formative feedback [29, 31] on students' work that otherwise might need to come from human interaction, but little is understood on how the availability and design of autograders change computing students' help-seeking landscape, and how students utilize automated feedback as signals to seek other forms of help, except for some recent studies that showed the design/availability of autograder feedback influenced students' usage of social resources [2, 12, 90].

Finally, **generative AI (GenAI)** tools and **large language models (LLMs)** have emerged as a new informal, highly accessible, and synchronous help resource [33, 63, 77]. However, students' adoption level varies [33, 63], and it is unclear whether students treat GenAI/LLMs as fundamentally different help resources from static online resources [77].

Comparison of Perception vs. Behavior. In addition to UTAs, the surge in class sizes has also mandated the adoption of educational technology to aid help-giving (e.g., CH queueing apps [60], online class forums [13, 65], and dedicated platforms for code autograding [9, 28, 88]). These platforms thus enable fine-grained data collection to examine students' *actual behavior* alongside their *self-reported usage pattern/decision process*. Such data has already been leveraged to understand the efficacy and/or design of *one* or *two* specific help resources at a time (see Section 3) but has yet to be used in advancing research on resource selection, as existing literature on resource selection mostly relied on students' self-reported *frequency* and *order* of resource usage [10, 14, 33, 89].

Lack of Emphasis on Sequentiality. Makara and Karabenick's *expectancy-value model* of resource selection [55] implicitly assumes that *every help episode* (i.e., each time a student needs help and self-identifies the need) *is independent of each other*, and that the student would *select one or more resource(s) all at once*. However, multiple works [10, 14, 89] have since revealed *sequentiality* in students' help-seeking approaches through interviews: They tend to first seek help from "informal albeit less useful resources," then consciously *transition* to "more formal and useful resources" until their help need is fulfilled. In between the extremes of sequentiality and simultaneity, few existing works have tackled the middle ground of *semi-sequential resource utilization*: how/why multiple help resources are used *in immediate succession* to complement each other, or how/why students utilize the same resource *repeatedly within the same episode*.

Motivated by the above gaps in existing literature, our work seeks to study computing students' *help-seeking resource selection and utilization* with awareness of *resource sequentiality*, using both self-reported and actual behavioral data, while the latter is limited to the course-affiliated resources. Our **research questions (RQs)** are:

(RQ1) When explicitly asked about the *order* of resource usage in their help-seeking approach, what patterns or progressions do students' answers exhibit, and how can we interpret them with existing help resource dimensions?
(RQ2) How much do students' actual utilization of formal help resources match their self-reported approaches, and where do mismatches happen?
(RQ3) What patterns emerge when students' help-seeking records are sequentially examined in finer granularities than an entire semester?

To answer RQ1, we surveyed students' self-reported *order* of help resource usage across seven courses at two institutions. Several clusters of help resources emerged from our results and formed a progression (Section 5.2). Through the lens of help resource dimensions, we found that no dimension alone can explain the entire progression, and that many dimensions pose tradeoffs to students in their resource selection (Section 5.3). Our results partially confirm existing phenomena reported in the literature, such as the *detached to anchored* progression found in Wirtz et al. [89] and Doebling and Kazerouni [14]'s works, but also show that they do not generalize to all resources. We also identified two help resources, GenAI/LLMs and People Unaffiliated, as more influenced by instructional context and course topics than the other resources (Section 5.4).

To answer RQ2 and RQ3, we analyzed students' help-seeking records from instructor/**graduate TA (GTA)** CH, UTA CH, class forums, and two types of autograder platforms from two of the seven courses across a total of 11 offerings. We reconstructed *help-seeking event sequences* at the student-assignment level. For RQ2, we found students' actual orders of help resource usage do not deviate much from their self-reported orders when limited to course-affiliated social help resources tracked by our data (Section 6.2), suggesting the findings in RQ1 may be sufficiently indicative of students' actual behavior. For RQ3, through mining frequent associate rules (of help-seeking events) in the sequences, we found social help-seeking led to measurable progress (as indicated by

Table 1. Dimensions in the Help Landscape, along with How They Were Described to Students in a Recent Study [46], Sorted in Order of Perceived Importance Reported by the Study

Dimension	References	Text description in recent study [46]
Timeliness	[26, 73, 82]	"How quickly I can get a response to my help request"
Availability	[26, 69, 73, 82]	"How often the help resource is available to me"
Adaptability	[55, 69, 82]	"How much the help resource can tailor its response to my specific request"
Time Anchor	[14, 89]	"Whether the help resource is only available at a specific time in the day/week"
Effort	[35, 73]	"How much effort it takes me to provide the resource with the context needed to understand and solve my problem"
Space Anchor	[14, 89]	"Whether the help resource is at a specific physical location"
Formality	[14, 17, 26, 55, 70, 89]	"Whether the help resource is officially part of our course"
Socialness	[10, 26, 55, 70, 73, 82]	"How much the help resource involves interacting with people"
Media richness	[70, 73]	(Not surveyed in the study [46])

Accessibility, mentioned in many works [14, 69, 82, 89], is an umbrella term that captures Availability, Time Anchor, and Space Anchor (see Section 5.2 for more discussions).

the autograder results) more often than not (Section 7.2). We found students' usage of CH itself was the best indicator for future usage of CH *within the lifespan of the same assignment*.

While limited by the scope of our data collection and analyses, our work serves as a preliminary attempt to rethink computing students' help-seeking process with an awareness of resource sequentiality, so as to better understand and foster effective help-seeking.

2 Theoretical Framework

In this section, we introduce existing theories on help-seeking (Section 2.1) and help resource selection (Section 2.2), the help landscape dimensions (Table 1), and a recent work on students' perceptions of the dimensions (Section 2.3).

2.1 Foundational Work in Help-Seeking

Help-seeking is widely considered a metacognitive and SRL strategy [38, 39, 53, 55] adopted by students to support their learning. Stites et al. [82] point out that students' help-seeking processes align with the *forethought* (identify the need and determine help resources to utilize), *performance* (solicit and obtain help from multiple resources), and *self-reflection* (process, judge, and react to help received) phases in Zimmerman's SRL model [95].

Karabenick and Dembo [39] outline eight (not necessarily sequential) *stages* of the help-seeking process: (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 summarized by the meta-analysis by Fong et al. [17], several complexities influence both the likelihood a post-secondary student seeks help and the type of help they seek. Different types of help form a spectrum spanning from *instrumental* to *executive* (or *expedient*): Instrumental help emphasizes acquiring skills (the "process"), which often also reduces the need for subsequent help, while executive/expedient help focuses on getting things done (the "outcome"), which only reduces the short-term workload of students. The meta-analysis confirms that instrumental help is beneficial to students' academic achievement indicators (such as GPA and test scores, with a larger effect on the former), whereas both executive help-seeking and avoidant help-seeking (i.e., refusal to seek help) are disadvantageous for learning. Other works [36, 37, 40] indicate that students with a focus on mastery (comparing oneself with oneself) are less avoidant, less likely to feel their self-esteem threatened by help-seeking, and more likely to seek instrumental help, whereas students concerned with their performance (comparing oneself with others) are more avoidant, more likely to perceive help-seeking as a threat, and more inclined to seek executive help.

2.2 Help Resource Selection in Post-Secondary Contexts

Makara and Karabenick [55] introduce a four-dimensional framework to contextualize students' *perceptions* of help resources: (1) *role* capturing the formality of the resource; (2) *relationship* capturing the mental closeness students *perceive* towards the help resource; (3) *channel* distinguishing whether the help resource is over a media (textbooks or video) or from a person; and (4) *adaptability* measuring the help resource's capability to provide personalized help. They also propose an expectancy-value model of resource selection as a refinement of stage 5 (*decide on whom to ask*) of Karabenick and Dembo's framework [39], theorizing a student's chances of seeking help from a particular resource as influenced by (1) the accessibility/availability of the resource, and (2) the student's *perceived* usefulness of the resource and whether their desired type of help (stage 4) match what the resource provides.

Makara and Karabenick's expectancy-value model has inspired a line of works studying students' resource selection process both in the general post-secondary education context [26, 34, 70, 73] and engineering/computing educational contexts [14, 33, 69, 82, 89]. This line of work collectively provides a set of *help resource dimensions* (see Table 1) that we explicitly use in our work to describe and analyze the help resources. Qayyum [70] finds students in cohort-modeled programs (i.e., students take most of their classes with the same classmates throughout the program) prefer seeking help from classmates (low Formality) rather than instructors (high Formality) and prefer in-person help-seeking (high Media richness) over other digital services. Stites et al. [82] cluster students using their usage frequencies of all resources, reporting a rich set of diverse usage frequency patterns that put different levels of emphasis on the Availability, Adaptability, Timeliness, Socialness, and *accessibility* aspects of help resources. Giblin et al. [26] point out that the classical works on help resources implicitly assume help comes from a list of predefined human resources (with high Socialness), while college students now have a wider menu of not only human resources but also technology-enabled non-human resources to seek help from. Their interviews mostly corroborate the elements in the expectancy-value model but also highlight Timeliness as a significant factor. On the other hand, Schlusche et al. [73] further focus on digitally mediated help based on the observation that social discomfort may prevent students from seeking face-to-face help with high Socialness. They explicitly argue for considering *synchronicity* as a factor in students' resource selection, and propose the "effort to phrase a problem via media" as a concern when seeking mediated asynchronous help.

Through an exploratory factor analysis on engineering students' self-reported usage frequencies and perceived usefulness of all help resources, Wirtz et al. [89] form *tiers* of resources of similar characteristics (in terms of the help resource dimensions). They identify an *order of usage* progression from low Socialness (*individualized* in their terms) to high Socialness resources. The article also proposes the concepts of Time Anchor and Space Anchor, where anchored resources are defined as "tied to a time or a space" (while *detached* resources are not). Wirtz et al. (and a subsequent replication study by Doebling and Kazerouni [14] in small-scale computing classes) reveal an *easily accessible, low utility -> not easily accessible, high utility* progression via student interviews. Among this line of works, the work of Price et al. [69] is the most domain-specific: They identify perceived Adaptability and accessibility as the key factors that influence novice programmers' willingness to seek help when programming.

On the front of help efficacy, a common focus is on the Formality dichotomy (sometimes also coined *internal/external* [77, 78]). By classifying all resources as either formal or informal, the meta-analysis by Fong et al. [17] finds only formal help-seeking to be significantly correlated with students' achievement, although it also acknowledges that formal resources (mostly instructors) generally give higher-quality instrumental help than informal resources, potentially confounding the result. UTAs are simultaneously formal (affiliated with the course) and informal (as peers) and thus do not truly belong to either end of the dichotomy; few of the contexts analyzed by Fong et al. rely on UTAs to provide the bulk of help like in some of today's large-scale computing courses [57, 59]. Therefore, this dichotomy has become a spectrum, while the effectiveness of "medium Formality" help provided by UTAs remains unclear; in fact, recent works have reported that UTAs tend to give executive help in large CS classes [47, 48].

2.3 Help Landscape

Our recent study [46] investigates how students value each dimension in the landscape (the first eight dimensions in Table 1, but not Media richness). The study finds students collectively value Timeliness as the most important among all dimensions, followed by Availability and Adaptability (in no particular order), then Time Anchor, Space Anchor, and Effort (in no particular order), while Formality and Socialness are deemed *least* important collectively by the students, despite being the two most studied dimensions in the literature. We organize the presentation (e.g., order of dimensions in Tables 1 and 4, and the discussions in Section 5.3) based on this order of importance.

3 Related Works

3.1 Resource-Specific Studies in Computing Contexts

Computing students now have access to more help resources than what has been studied in the general post-secondary educational context [33, 77], and many recent works in computing education have been devoted to studying these new help resources. We summarize them by whether or not the help resource(s) being studied comes from other people/humans. Note again we use the term CH to refer to what past works also call office hours or tutoring hours.

3.1.1 High Socialness Human Help Resources. For seeking social help from a human, this includes in-person [79] and online course-specific CH [23], drop-in help centers serving multiple courses [6], in-class [8, 74, 80, 84, 86] and publicly available [16] discussion forums, video calling platforms [5], and peer-formed study groups [66]. Existing works on these help resources report benchmark statistics such as usage rates of class forums [76], wait time, interaction length, and the number of visits per student in their CH [23, 25, 44, 45], while also categorizing the kind of help students seek from these resources [21, 25, 44, 45, 71, 83, 86, 87].

Past works report prior programming experience [75, 93], major [14], academic confidence [93], modality of help [23], and affective experiences [51] as potential factors that influence the usage of human help. While several works [8, 86] find higher-performing students actively seek help online on the class forums more than their lower-performing peers, Smith et al. [80] report viewing more questions/answers on the forums to have a larger effect on student performance than asking questions, writing answers, or interacting with instructors.

It has been reported that women seek help more frequently than men in CH [14, 23, 93], on class forums [74, 84], and from online resources like StackOverflow [81] despite being underrepresented [85], and are willing to wait longer in queue for CH help [22]. Recent works also report race and intersectionality [74] and prior experience [75] influence whether students choose to remain anonymous when posting on class forums.

In terms of usage behavior, it is found that students usually provide insufficient context when asking questions in CH [20, 25] and on class forums [20, 86]. Timeline analysis on CH, class forums, and video calling platforms [5, 20, 44] shows students' help-seeking behavior peaks immediately before assignment deadlines, and the type of help students seek differs between peak and non-peak time.

3.1.2 Low Socialness Non-Human Help Resources. Static Online Resources had been a viable low Formality (as they are external to the course) help resource to computing students [78] long before the emergence of GenAI and LLM transformed the learning experience. Earlier studies around the time GenAI/LLMs tools emerged either set up GenAI tools (such as OpenAI Codex [42, 62] and GitHub Copilot [27, 67]) in controlled environments for novice programmers to interact as a learning tool, or compare the quality of help given by different LLMs to novice programmers' help requests in hypothetical situations [32].

More recent works have started to study GenAI/LLMs as an available help resource that students can self-choose to interact with, either as a standalone external tool [33, 63, 77] or pedagogically designed and wrapped as ChatBots [91]. These works all report a varied level of adoption and reliance on GenAI/LLMs among students that correlate with self-efficacy and performance [63]. It is unclear whether students view GenAI/LLMs as a fundamentally different help resource from Static Online Resources [77].

Within the classroom, computing courses have incorporated automated feedback in **learning management systems (LMS)** [11] and specifically autograders [43] to scale instruction. Autograders situate in a unique place in computing students' help-seeking landscape: Not only are they *field-specific* (and thus less studied in help-seeking literature from the general educational contexts), but they are also a *passive help resource*: Students submit their work through the autograders and get the automated feedback regardless of their intent to seek help. In other words, students can skip the first six stages in Karabenick's help-seeking process (Section 2.1) and still get helped by the autograders. In the computing education community, much effort has been spent on studying giving more adaptive [58] or formative [29, 31] feedback in those tools. However, they are so synchronous and accessible that recent works have revealed students' over-reliance on autograders [1] and investigated whether slightly delayed feedback is more constructive [50].

3.1.3 Interaction between Help Resources. Although the above works are situated around computing classrooms, most of them only study one or two specific help resources at a time and within a specific instructional context, rather than directly investigating computing students' help-seeking process or resource selection approaches.

In contrast, relatively few works have attempted to detangle how help resources *interact* with each other. Although some recent work specifically mentions this gap of knowledge [8], most of the analyses therein still focus on each resource separately. A cross-course analysis of annotated student questions in CS1/CS2 class forums by Mannekote et al. [56] reveals students' class forum usage correlates with the availability of alternative help resources such as CH. Deorio and Keefer [12] report that the design of autograders (e.g., whether test cases are visible) influences the amount of help sought in CH, showing that human and non-human resources are not separate parts of students' help-seeking landscape. Gao et al. [21] find that a majority of students who seek help in CH make measurable progress in autograders within 20 minutes of their CH interaction. Two other works [2, 90] study how autograders shape student help-seeking behavior in introductory programming courses. While Basu et al. [2] show the availability of autograders can reduce the number of questions students ask on the class forum, Wrenn and Krishnamurthi [90] report that many questions students asked in class forums could have been answered automatically by the available autograder, implying that students did not fully understand how to leverage the power of their autograder.

Table 2. Overview of All Courses in Our Data

Inst.	Course	Shorthand	Enrollment	Student body	Type	Prerequisite
NCSU	Intro. to computing in MATLAB	MATLAB	294	Non-majors/engineering	Service	-
Duke	Intro. to programming in Python	PYTHON	171–262	Non-majors/undeclared	Service/Entry	-
	Data structures and algorithms	DSA	177–272	Majors	Required	PYTHON or equivalent
	Data science	DATA	66–160	Interdisciplinary	Popular elective	PYTHON + (DSA or statistics)
	Discrete math	DM	120–138	Majors	Required	DSA
	Database systems	DB	330	Majors	Popular elective	DSA + a systems course
	Algorithms design/analysis	ALGO	123–220	Majors	Required	DSA + DM

See Table A1 in Appendix Section A.1 for detailed enrollments and demographics.
NCSU, North Carolina State University.

3.2 Mining Sequential Educational Data

There has been ample work on **educational process mining (EPM)** [3] using event logs collected from LMSs [7, 15, 61], Massive Open Online Courses [4, 54, 72], and **intelligent tutoring systems (ITSs)** [41]. Similar to all EPM works, we investigate patterns emerging from students' event-based log data. However, our RQ3 (and therefore its data collection and analyses) has a much narrower scope: We focus on computing students' sequential *help-seeking* records in course-affiliated resources to understand their resource usage and utilization patterns. In contrast, most EPM works investigate the entire learning process, where help-seeking action is either not recognized or limited to help features native to the learning system (e.g., hints in ITSs [41]).

4 Contexts and Data Collection

4.1 Instructional Contexts and Participants

We collected data at **North Carolina State University (NCSU)**, a large-size public university, and Duke University (Duke), a medium-size private university. Both institutions are research-oriented, located in the United States, and use 15-week semesters. Data was collected from a total of seven courses (see Table 2 for an overview), spanning from Fall 2021 (Fa21) to Spring 2024 (Sp24), with varied types of datasets collected from each offering (see Table 3 for a summary). Throughout this article, we refer to the seven courses (rows in Table 2) as *courses*, and single offerings of a course (the 19 *cells* in Table 3) as *classes*. We report the demographic distributions of all classes in Table A1 in Appendix Section A.1.

4.2 The Order Survey of Resource Usage

4.2.1 Survey Instrument. Starting from Sp23, we asked all students to group and rank all available help resources (listed in Table 4) in their ideal *order of resource usage*. 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. This wording was designed specifically to obtain students' *availability-agnostic* order of resource usage. Students could use up to three groups and could omit resources that they did not use. We hereon refer to this survey as "the Order survey."

The Order survey, with a focus on students' *order* of using the resources, is most similar to the interviews in recent works [10, 14, 89] that revealed sequentiality in students' help-seeking habits, while fundamentally different from asking students to report their *frequencies* of using each of the resources (quantitative parts in [14, 89], among others [33, 82]). The former studies students' typical behavior within a single help *episode* [55] (i.e., from identifying the need of help, deciding

Table 3. Summary of Data Collection Scopes

Inst.	Course	Fa21	Sp22	Fa22	Sp23	Fa23	Sp24
NCSU	Matlab						○
Duke	Python	×	×	×	×	⊗	⊗
	Data		×	×	×	⊗	⊗
	DSA					○	○
	DM					○	○
	DB					○	
	Algo					○	○

Classes with a ○ in their cells (where the Order survey was administered) are used in RQ1. Classes with a × (where behavioral data from CH, class forums, and autograders were all collected at the assignment level) are used in RQ3; see Table A2 in Appendix Section A.2 for more contexts of course-affiliated help resources in these classes. Classes with ⊗ (where both kinds of data were available) are used in RQ2.

Table 4. Help Resources Included in the Order Survey, via the Perspective of Help-Seeking Dimensions

Resource name	Timeliness	Availability	Adaptability	Time anchor	Effort	Space anchor	Formality	Socialness	Media richness
Class Material	n/a	H	L	L	n/a	L	H	L	L
Static Online Resources	n/a	H	L	L	n/a	L	L	L	?
GenAI/LLMs	H	H	M	L	M	L	L	M	L
Classmates	?	?	?	M	?	?	L	H	?
Forum-read	n/a	H	L	L	n/a	L	M	M	L
Forum-write[a]	L	H	H	L	H	L	M	M	L
UTA CH in-person[a]	H	M	H	H	L	H	M	H	H
UTA CH online[a]	H	M	H	H	L	L	M	H	M
GTA+ CH In-person[a]	H	L	H	H	L	H	H	H	H
GTA+ CH online[a]	H	L	H	H	L	L	H	H	M
People unaffiliated	?	?	?	?	?	?	L	H	?
Autograder feedback[a] (not surveyed)	H	H	H	L	n/a	L	H	L	L

We describe the resources as H (*high*), M (*medium*), L (*low*), ? (*depends*), or not applicable, along each dimension. Dimensions are ordered by their relative importance reported in the recent study [46], in which Media richness was not surveyed. GTA+ stands for *graduate TA and instructors*. The pilot survey in Sp23 did not distinguish between Static Online Resources and GenAI/LLMs. Depending on each class's help context, some options may be omitted or merged.
[a]Resources bolded are captured by the assignment-level behavioral data collection.

to seek help, to determining resource(s) to seek help from), whereas the usage frequency-based surveys study students' (average) behavior across multiple episodes, without an emphasis on the sequentiality in students' help-seeking. The Order survey provides the basis of our analyses of RQ1 while also facilitating that of RQ2 (by comparison with behavioral data in select classes).

In the Order survey, we differentiated between *passively* reading existing threads on the class forum (called Forum-read) and *actively* posting new threads/comments to seek help (Forum-write). This is because past works have characterized the behavior of *class forum lurkers* [93] (i.e., students that mostly read others' posts but not post their own) and called for an explicit differentiation between passive and active help-seeking behaviors [94].

We also differentiated between in-person and online CH as existing works have identified a difference between students' usage of CH in different modalities [23]. In classes that offered *hybrid*

Table 5. Summary of Behavioral Datasets by Course

Course	CH (# of interactions)				Forums (# of contributions)						Autograders (# of attempts)		Sequences constructed
	Raw	Consenting	Valid	Assignment	Raw	Consenting	Help-seeking		Assignments		Homework	Project	
							Student	Staff	Student	Staff			
Python	4,288	3,407 (79.5%)	2,965	2,564	12,346	9,332 (75.6%)	4,055	3,360	1,912	1,114	120,893	38,812	10,550
Data	1,675	1,158 (69.1%)	1,018	889	7,900	6,199 (78.5%)	1,782	1,882	760	635	39,077	-	5,560

See Table A3 in Appendix Section A.2 for details per class. Numbers in parentheses are proportions over raw data counts. Failed autograder runs (e.g., submissions with wrong files or server technical issues) are omitted from the counts. We do not report raw data counts for autograder datasets as the raw datasets contain many entries from course staff for testing. Forum contributions do not include edits to existing content.

CH (i.e., TAs helping students in person and online *simultaneously* during a shift), the resources were phrased as *attending CH in person* and *attending CH online*. This captured the modality of the student (and thus the help interaction) and de-emphasized the types of modalities offered by the help-giver.

Initially in Sp23 as a pilot, the survey used “online resources” to capture both static and dynamic online help resources external to the course. In response to the rise of GenAI tools and LLMs, we refined the survey from Fa23 to distinguish between these two kinds of resources as they are quite distinct in multiple dimensions in the landscape (see Table 4).

Not all classes had autograders (e.g., DM typically involves no programming at Duke). In classes that used autograders, since students submitted their work through autograders, and thus got automated feedback from the autograders regardless of their intent to seek help, it did not make sense to include the autograders as an option in our survey.

4.3 Behavioral Data Collection in Internal Help Platforms

We collected from Python and Data students’ usage records in all course-affiliated help resources (CH, class forums, and autograders) *at the assignment level*, detailed below.

4.3.1 CH. Most (but not all) of the regular CH in all classes were offered in the evening using the MyDigitalHand [60, 79] queueing app. Students sought help by submitting a request on MyDigitalHand regardless of modality. Before submitting the request, students indicated what they were working on, usually but not always at the *assignment*-level (e.g., “Project 2”). We also collected the timestamps of each request and the corresponding interaction. The scope of MyDigitalHand usage, and thus the coverage of the data, includes UTAs in every class and GTAs/instructors (GTA+ as a shorthand) in some classes (see Table A2 in Appendix Section A.2).

The sizes of the datasets are shown in Tables 5 (per course) and A.3 (per class, in Appendix Section A.2). A total of 79.5% of Python and 69.1% of Data interactions remained after filtering for consent. Following existing works [23, 44, 45], we further excluded the interactions where the waiting time (the time student stayed in the queue) exceeded 4 hours or the interaction length was not between 1 and 60 minutes. These may represent human errors (e.g., students not showing up after submitting a request, or TAs not closing an interaction at the end of a shift). We excluded these interactions as they might represent help-seeking actions that did not actually take place. The remaining interactions were considered *valid*. A total of 86.4% of Python and 87.3% of Data valid interactions were annotated with an assignment; the others were likely related to exams, general course content, or technical issues. Note that the number of available hours (see Table A2) and available TAs per shift (not shown) varied across classes, potentially impacting the dataset sizes.

4.3.2 Class Forums. All classes used Ed Discussion [13] as the class forum (just *forum* in later parts of the article). Datasets exported from the platforms were preprocessed into tables of *contributions*, where each contribution represented a user (student or staff member) posting some text to the forum, including not only new threads but also responses to existing threads. A total of 75.6% of Python and 78.5% of Data contributions remained after filtering for consent. We then further excluded all threads initiated by a staff member: Some classes actively used the forum in their class meetings, and therefore the staff-initiated threads mostly represented live class discussions instead of help-seeking attempts. The remaining interactions are called *help-seeking* contributions in Table 5 and were further separated into *student contributions* (including students' initial posts/questions, students' follow-ups, and students' responses to other students' threads) and the *staff responses/comments*.[1] Similar to CH interactions, not all contributions were associated with an assignment. When available, we used students' self-annotations of the threads to identify the assignment that each contribution was associated with. Otherwise, we parsed the title, text, or timestamps to infer the assignment (see Appendix Section A.2 for the details and a summary of the attributes of contributions used in this work).

4.3.3 Autograder Logs for Programming Assignments. Both Python and Data utilized Python-based formative programming assignments (Table A2 in Appendix Section A.2). We excluded summative assignments from our work because students were prohibited from seeking help on them. We also excluded optional assignments due to the associated selection bias and the sparsity of corresponding help-seeking data.

With a focus on basic programming, Python administered two types of formative programming assignments:

— *Homeworks*: Each homework consisted of several independent "write a function that"-type questions with well-defined input and output formats.
— *Projects* that involved writing multiple interconnected functions and files at a time.

Data had weekly homeworks focusing on applying libraries to conduct data analyses. Each Data homework contained multiple (mostly independent) questions that were meant to be solvable with 10 lines of code or less.

The Python homeworks were administered on an internal server and contained no manually graded parts. Python projects and Data homeworks were administered on Gradescope [28], where most (but not all) of them also included a manually graded component. For each autograder attempt of each assignment, we collected the timestamp, the score of the attempt, and the maximum possible score. Although the autograder logs were mostly at the *question* level, all analyses in this work were conducted at the *assignment* level because the CH and forum datasets were collected only at the assignment level. The sizes of the datasets are included in Tables 5 and A3.

4.3.4 Help Sequence Construction. For each student in these classes and for each assignment, we linked all datasets to construct a chronological *sequence of events*, where an event represents the student:

— submitted a CH request (i.e., "entering the queue");
— started a CH interaction with a staff member;
— ended a CH interaction;
— initiated a new thread on the forum;
— received a response (answer/comment) from others on the forum;
— self-commented on an existing thread initiated by themself on the forum;

[1] The size of *staff responses/comments* is sometimes larger than that of *student contributions* because both are filtered by consent and a question/post may trigger multiple responses.

- attempted the autograder *with progress* (i.e., achieving a personal new high score for the assignment in question);
- attempted the autograder *without progress* (i.e., getting less than or equal to current personal high score);
- attempted the autograder and *passed all tests*.[2]

Once the student had passed all tests in the autograder, the autograder could no longer provide any more meaningful help, and thus any additional attempts only implied the student continued working on the manually graded parts of the assignment. Therefore, all autograder attempts after an "All Tests Passed" event were discarded from the sequence. We chose not to track the following events because they are rare, invalid, or not part of students' help-seeking behavior:

- CH interactions with invalid timestamps (12.7% of consenting interactions; see Section 4.3.1 for the definition);
- student answering others' forum threads (more likely to be help-giving rather than help-seeking);
- failed executions of the autograder (0.1%–1.5% of the raw autograder datasets);[3]
- staff submitting to the autograder on behalf of the student (<0.1% of all datasets).

The number of sequences constructed can be found in Table 5 (per course) and Table A3 (per class, in Appendix Section A.2).

Table 6 demonstrates an example of a Data student's sequence of 21 events for a homework. During a 3-day timespan, this student submitted nine times to the autograder (six with progress), passing all tests in the final attempt. The student also posted three question threads on the forum (event IDs 2, 9, and 14) and used CH once (event IDs 17–19) to seek help. Note that the last event (event ID 21) was a late (33 hours) response by a staff member to the student's follow-up question (event ID 11) on the second question thread. That question was likely resolved in the CH interaction as the student passed all tests in their final autograder attempt (event ID 20) that took place right after the interaction (event ID 19).

5 RQ1: Students' Help Resource Usage Order

5.1 Metrics: Borda Score and Dominance

To enable characterization of the sequentiality of students' help resource usage, we analyzed the Order survey results with the Borda Score and Dominance metrics.

Borda Score. Borda Scores [92] are used extensively in social choice and voting theory to quantify ordinal data. For our context, we assigned each response that put a specific help resource into the student's *used first*, *used second*, *used third*, and *omitted (would never use)* groups a Borda Score of 3, 2, 1, and 0, respectively. The Borda Score for each resource within a class is then the mean Borda Score over all students in that class.

Dominance. The Borda Score captures each resource's position in students' ideal order of resource usage *relative to all other resources* without any specific focus on comparing the resource with one specific other resource. We capture the *pairwise* relations using Dominance: For each *pair* of

[2]This information was only available from the Gradescope-administered assignments. For the Python homework internal server, the autograder for each question was independent, and some homeworks only required completing a fixed amount of questions out of all provided. Therefore, we could not infer the "all tests passed" status from aggregating the question-level autograders.
[3]This mainly included autograder submissions in which students submitted the wrong files or selected the wrong question to submit to.

Table 6. An Example of a Student's Event Sequence in a DATA Homework

Event ID	Event code	Time to deadline	Autograder score
1	Autograder (progress)	T-3 days 09:36:30	9/56
2	Forum thread initiation	T-3 days 09:15:18	
3	Forum self-comment	T-3 days 08:38:06	
4	Autograder (progress)	T-3 days 08:34:44	21/56
5	Forum response	T-3 days 08:28:30	
6	Autograder (progress)	T-2 days 02:03:58	32/56
7	Autograder (progress)	T-2 days 01:43:58	42/56
8	Autograder (no progress)	T-2 days 01:32:12	42/56
9	Forum thread initiation	T-2 days 01:25:44	
10	Forum response	T-1 days 23:48:18	
11	Forum self-comment	T-1 days 23:12:30	
12	Autograder (no progress)	T-1 days 10:41:22	42/56
13	Autograder (no progress)	T-1 days 10:40:03	42/56
14	Forum thread initiation	T-1 days 08:55:08	
15	Forum response	T-1 days 04:48:35	
16	Autograder (progress)	T-1 days 01:33:16	48/56
17	CH request	T-1 days 01:31:15	
18	CH start	T-1 days 01:29:55	
19	CH end	T-1 days 01:08:59	
20	Autograder (all tests passed)	T-1 days 01:07:56	56/56
21	Forum response	T-0 days 14:16:22	

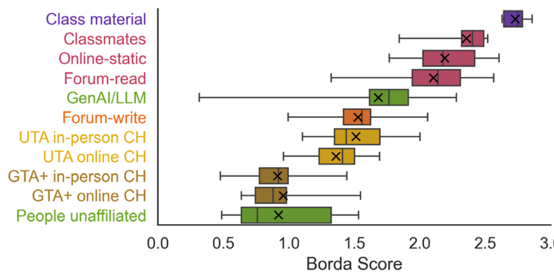
resources A and B, the Dominance of resource A *over* resource B within a class is the percentage of students that ranked resource A *strictly before* resource B in their responses to the Order survey.[4]

Figure 1 shows the results of the Order survey *aggregated across all classes*. Specifically, Figure 1(a) plots the distributions of the class-specific Borda Score of all resources, and Figure 1(b) is a heatmap of the *mean* Dominance across all classes, where each cell shows the (rounded) mean Dominance of the *row* resource over the *column* resource. For example, 81% of the students in a typical class (top-right cell in Figure 1(b)) indicated that they would consult Class Material strictly before seeking help from People Unaffiliated, whereas only 4% (bottom-left) said the opposite. Note that symmetric cells on both sides of the diagonal generally do not add up to 100, as some students put both resources into the same group (as can be inferred, 15% of students did so for Class Material and People Unaffiliated).

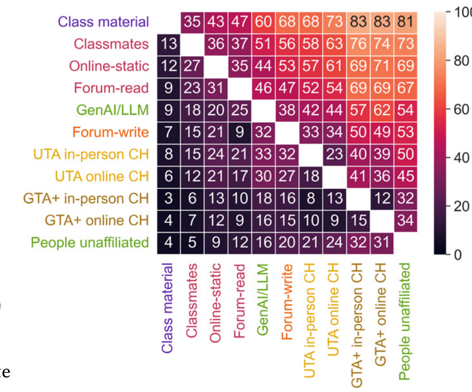
5.2 Progression of Help Resources

As shown by the coloring (of the resources) in Figure 1, both the Borda Score and Dominance results reveal a progression of several *clusters* of resources akin to the *tiers* in the work by Wirtz et al. [89]. Intuitively, each cluster contains a set of resources that not only share similar properties (regarding the landscape) but also are used by students in similar ways. Students mostly reported

[4]Note that Dominance does not take into account how far apart a student separates the two resources in their own order. That is, for a student that put resource A into their *used first* group, whether resource B was put into *used second*, *used third*, or *omitted* has no impact on the Dominance of resource A over resource B, since each of these three options are strictly after *used first*.



(a) Box plot of Borda Score distributions. Whiskers show the complete data range. Resources are ranked by median Borda Score. Each color represents a resource cluster discussed in Sections 5.2 and 5.4.



(b) Heatmap of mean Dominance for each resource pair.

Fig. 1. The Order survey results aggregated across all classes.

using resources in *earlier* clusters strictly before seeking help from those in *later* clusters across classes, while the order of resources within the same cluster fluctuated.

However, not all clusters are equally consistent at the *class* level (see Figure B1 in Appendix B for the per-class results). Although the Order survey attempted to reveal students' *availability-agnostic* order of resource usage, class design and policies still significantly influenced how students interact with help resources, especially for newly emerged ones. Among all resources, two (GenAI/LLMs and People Unaffiliated) seem to be highly context-dependent, as evidenced by their wide total/interquartile ranges in Figure 1(a).

The progression of clusters[5] with GenAI/LLMs and People Unaffiliated removed is:

(1) Class Material
(2) Classmates, Static Online Resources, and Forum-read
(3) Forum-write
(4) UTA CH (including in-person and online)
(5) GTA+ CH (including in-person and online)

In the next subsection, we interpret the progression above using the dimensions in the help landscape, comparing and contrasting with what past works reported. We then come back to discuss the two excluded resources in Section 5.4.

5.3 Interpreting the Progression via the Help Landscape

In the following, we discuss the dimensions in order of their relative *perceived order of importance* found in our recent study [46]. We hypothesized that dimensions deemed more important by the students are more likely to be able to truly reflect the sequentiality in students' help-seeking approaches.

5.3.1 Timeliness and Availability. Timeliness ("how quickly I can get a response to my help request") and Availability ("how often the help resource is available to me") are respectively the

[5]This progression is consistent with *most but not all* of the per-class results. Most of the discrepancies see Forum-write getting a lower Borda Score than UTA CH In-person (5 classes) or both kinds of UTA CH (2 classes). While the sequentiality between Forum-write and UTA CH is less pronounced than between other pairs of clusters, we opted to keep them as separate clusters as they have drastically different characteristics regarding the help landscape (see Table 4 and the discussions in Section 5.3 below).

first and second most important dimensions students consider both for *what resources to use* and for *what resources to use before others* [46]. However, even among only the resources for which Timeliness applies, the clusters found in Section 5.2 do not simply follow the Timeliness dimension; for example, students reported using cluster 3 (Forum-write, low timeliness) earlier than UTA CH and GTA+ CH, both with high timeliness. On the other hand, although the clusters do roughly progress from high to low Availability, the Availability dimension alone does not explain the sequentiality among clusters 1–3 (all high Availability except maybe Classmates). Instead, these two important dimensions *together* better explain the progression: Among the high Availability resources in clusters 1–3, students are more likely to utilize the high Timeliness ones (Class Material, Static Online Resources, and Forum-read) before the low Timeliness ones (Forum-write, and maybe Classmates).

5.3.2 Adaptability and Effort. Adaptability is reported as the third most important dimension in the landscape [46]. As defined in Makara and Karabenick's work [55], the Adaptability of a help resource characterizes how much the resource is capable of personalizing the help towards the help request (e.g., a textbook is completely nonadaptive, whereas a personal tutor is ideally very adaptive). Despite being one of the four dimensions proposed by Makara and Karabenick to characterize help resources, the nonadaptive → adaptive progression has been less explicitly studied in recent works on student's help resource selection.

We argue that the Adaptability and Effort dimensions present a natural tradeoff in students' resource selection process: To seek highly adaptive (preferable) help, students need to *actively* describe/phrase the contexts of their help requests rather than *passively* receive information [94], which in turn requires high Effort (not preferable). This tension can be seen in our data: Although Forum-read (low Adaptability and Effort) has a higher Borda Score than Forum-write (high Adaptability and Effort) in *all classes* (likely due to Availability reasons), this distinction is more pronounced in more advanced classes than introductory, programming-heavy classes (Python, Matlab, and DSA),[6] likely due to the Effort overhead in advanced classes being relatively more costly.

5.3.3 Time Anchor, Space Anchor, and Media Richness. Per the definition by Wirtz et al. [89], help resources are *temporally anchored* if "tied to a time" and *spatially anchored* if "tied to a space," and students follow a *detached* → *anchored* progression. While temporally anchored resources are inherently timely, *not all timely resources are temporally anchored* (e.g., autograders and GenAI/LLMs are usually timely but temporally detached). Therefore, Time Anchor is conceptually different from Timeliness. Besides Classmates, whose levels of Time Anchor or Timeliness are context-related and hard to measure, our students reported using temporally detached (i.e., low Time Anchor) resources (Class Material, Static Online Resources, Forum-read, and Forum-write) before any kind of CH with high Time Anchor.

However, most of our results do not agree with the *detached* → *anchored* progression regarding spatial anchor: Both kinds of in-person CH (UTA CH Online and GTA+ CH Online) have a slightly higher mean Borda Score than their online counterpart (UTA CH in-person and GTA+ CH in-person, respectively) in Figure 1(a), while at the class level, 11 of the 18 pairs of in-person vs. online CH offered by the same kind of course staff see students reporting to utilize the in-person component first (see Appendix B). This result shows that *when both modalities are provided* (unlike works that compare exclusively in-person instructional contexts with pandemic-era exclusively online ones

[6]We computed the Borda Score difference between Forum-read and Forum-write for each class. The advanced classes have a median of 0.7 and a mean of 0.65, whereas the introductory classes have a median of 0.47 and a mean of 0.48. Furthermore, four of the five classes with the smallest Borda Score differences between Forum-read and Forum-write are introductory classes.

[23]), the Space Anchor and Media richness dimensions pose another tradeoff (as in-person CH is media-richer than online ones) to the students.

While some past works unified the Availability, Time Anchor, and Space Anchor dimensions by the concept of *accessibility* [26, 35, 69, 73], our nuances show that they should be treated as separate dimensions.

5.3.4 Formality. The *informal* → *formal* progression is reported as an emergent theme in Doebling and Kazerouni's interviews [14]: Students tend to seek help from less formal resources first before resorting to more formal ones, *despite perceiving the formal ones to be more useful.* More specifically, they report it as the *online resources* → *peers* → *instructor (asynchronous)* → *instructor (synchronous)* progression, where the concept of peers seems to *not* include TAs as near-peers (Section 8.4 therein [14]: "participants did not talk about TAs often"). That specific progression roughly maps to Static Online Resources → Classmates → Forum-write → GTA+ CH in our context. This is echoed in our results, although Static Online Resources and Classmates appear in the same cluster; in fact, 8 of the 12 classes (see Appendix B) reported seeking help from Classmates earlier in aggregate.

However, the *informal* → *formal* progression does not completely generalize to all resources in our data, nor do students perceive Formality as an important dimension [46]. Class Material, by definition a highly formal resource, is in the first tier alone, predominantly used first by a strong majority of students. As shown in Figure 1(b), it even *dominates* all other resources (no other resources are ranked before Class Material by more than 13% of the students in any class).[7] It is likely that other dimensions such as Adaptability, Availability, and/or Time Anchor have contributed to Doebling and Kazerouni's progression.

5.3.5 Socialness. Wirtz et al. [89] report that among *strictly detached* resources (Class Material, Static Online Resources, Forum-read, and Forum-write), students normally seek help from the non-social (*individualized* in their work) resources before resorting to the social ones that require human interaction. This is mostly echoed in our data, as both Class Material and Static Online Resources precede both kinds of forum actions. However, this insight again does not generalize to all resources: Seeking help from Classmates is decidedly more social than Forum-write, but the former is mostly used earlier (56%) than the latter (15%) as shown in Figure 1(b).

5.4 Context-Dependent Resources

5.4.1 GenAI/LLMs. Strictly from the perspective of the landscape, GenAI/LLMs is most similar to—but seemingly even more favorable than—the other resources in cluster 2, as it is strictly more available than Classmates, strictly more timely than Forum-read, and strictly more adaptive than Static Online Resources. However, in aggregate, our results in both Borda Score and Dominance indicate that GenAI/LLMs is not prioritized by students more than the other resources in cluster 2; in fact, its Borda Score is lower than *all* resources in cluster 2 in 10 out of the 12 classes (see Figure B1 in Appendix B).

We attribute this to students' reluctance to adopt GenAI/LLMs as a frequent help resource. First, students would not be inclined to use—or at least *self-report using*—GenAI/LLMs if the course policies prohibit or advise against using them, and its relatively wide range of Borda Score (see Figure B1) may have reflected the influence of course policy on students' usage of GenAI/LLMs. In fact, MATLAB-Sp24, the only class that explicitly discouraged using GenAI/LLMs, is also the only class in which GenAI/LLMs ranks dead last.

[7] Note that Class Material is not a help resource in Doebling and Kazerouni's interview. Therefore, our results do not contradict theirs.

Even within a class where using GenAI/LLMs is allowed or encouraged, some students are still reluctant to adopt them as a help resource at all [33, 63] or engage with its feedback [91]. With the benefits and harms of leveraging GenAI/LLMs still a hot debate within the computing education community [49, 68], the attitude instructors pose towards the usage of GenAI/LLMs will likely continue to vary, rendering GenAI/LLMs a rather unstable help resource in the landscape in the foreseeable future [77].

5.4.2 People Unaffiliated. The relative position of People Unaffiliated in the progression varied by a wide margin, as it ranked as high as 5th (in class-level Borda Score, see Figure B1 in Appendix B) and as low as the last (in more than half of the classes surveyed). We note here that People Unaffiliated can include *upper-level students who have taken the course* as well as *people not in the same institution*. We believe this phenomenon can be broadly attributed to the Availability dimension: The fluctuation in students' usage of People Unaffiliated observed in our data reflects *how often students can access someone in their immediate social circle with adequate expertise in the course topic*. Such a person is relatively hard to find for advanced courses (DB and ALGO) or interdisciplinary courses (DATA), whose students rarely used People Unaffiliated, than introductory, programming-heavy courses (PYTHON, MATLAB, and DSA) whose students sought help from People Unaffiliated even earlier than TAs in aggregate.

5.5 Synthesis

By analyzing the data via Borda Score and Dominance, we revealed (1) several clusters of help resources that are similar in their characteristics and used in similar fashions by students, and (2) two highly context-dependent help resources (GenAI/LLMs and People Unaffiliated) whose usage fluctuated from class to class. We saw evidence suggesting *all dimensions* in the landscape factor into students' order of resource usage, while *no dimension alone can explain the entire progression*. Furthermore, our results supported the finding in our recent study on perceived importance of help landscape dimensions [46], as the dimensions deemed more important in that study (Timeliness, Availability, and Adaptability) are more compatible with the observed progression of clusters/resources than the dimensions that students self-reported as less important (Socialness and Formality).

In all, our results showed that (1) all dimensions in the help-seeking landscape need to be taken into account holistically to reason about students' order of resource usage, and (2) instructional contexts, such as availability of resources and course policies, affect students' decision process directly or indirectly.

6 RQ2: Students' Actual Resource Utilization Order vs. Self-Reported Approaches

We next turned to the student-assignment-level sequences to investigate whether students' actual order of help resource usage matched what they reported in the Order survey. On this front, we were naturally limited to comparing among course-affiliated resources (see Table 4) that allow for data collection, as we did not have data on students' utilization of external resources. Furthermore, we needed to exclude autograder feedback from this analysis, as autograders were utilized by the students both as an assessment tool (more often) and a help resource (less often) and were thus omitted from the survey.

As such, we focused on auditing students' utilization of resources among the Forum-write, UTA CH, and GTA+ CH clusters. The data used in this section includes all student-assignment-level sequences in PYTHON and DATA from Sp23 to Sp24 (i.e., the semesters that the Order survey was

Table 7. Student-Assignment-Level Sequences Used for Audit

Course-semester		# Consenting students	# Sequences Total	# Sequences Help-seeking	# Follows	# Reversals Strict	# Reversals Weak
PYTHON	Sp23	163	2,092	292 (14.0%)	321	30	20
	Fa23	189	1,114	310 (27.8%)	94	16	2
	Sp24	71	422	86 (20.4%)	35	11	6
DATA	Sp23	159	1,429	247 (17.3%)	265	22	24
	Fa23	65	585	58 (9.9%)	62	0	11
	Sp24	101	893	106 (11.9%)	102	10	7
Total		748	6,535	1,099 (16.8%)	879	89	70

Help-seeking sequences are sequences in which the student used Forum-write, UTA CH, or GTA+ CH. Note that follows and reversals are with regard to pairwise orders between clusters, and thus a sequence can contribute to multiple follows and/or reversals at the same time.

administered).[8] Only 16.8% (see Table 7) of these sequences per class are help-seeking sequences where the students sought active formal help in one of the three clusters.

6.1 Cases

For every student-assignment-level sequence, we examined whether the sequence showed the student *following* or *reversing* their self-reported pairwise order of resource cluster usage. Specifically, if the student indicated that they would utilize UTA CH (both in-person and online) strictly before Forum-write, their sequence was deemed a:

(a) *follow*, if UTA CH appeared before *any* Forum-write events.
(b) *strict reversal*, if a Forum-write event happened *during a time when UTA CH (either in-person or online) was available* before any UTA CH event. This is because the student could have accessed UTA CH *at the moment* while they were seeking help.
(c) *weak reversal*, if a Forum-write event happened *during a time when UTA CH was unavailable* before any UTA CH event. Since our survey asked the students to report their order of resource usage *assuming all resources were available*, the student here technically did not deviate from their self-reported order, but we also observed that they *did not wait until UTA CH became available* (there is usually some availability each day), resulting in the actual order of resource usage being different from their ideal order of resource usage.

Table 8 defines the subcases for each pairwise order of resource clusters (with the example above being Case 1), where subcases a, b, and c for each case represent follows, strict reversals, and weak reversals respectively.

6.2 Results

Figure 2 illustrates a case-by-case breakdown aggregated across all six classes in our audit analyses, where the result revealed significantly more follows than both strict reversals and weak reversals. For each pairwise order of resource clusters, either the sample size is extremely small (Cases 3 and 6, as few students ranked GTA+ CH earlier than the two other clusters), or there were way more follows than reversals of any kind.

[8]Note that we used both the pilot (Sp23) and revised (Fa23-Sp24) versions of the Order survey here. This is because the difference between the two versions only lied in whether they differentiated Static Online Resources and GenAI/LLMs, which is irrelevant to the analyses in this section.

Table 8. Definitions of follows and reversals for All Pairwise Orders among Forum-Write, UTA CH, and GTA+ CH

No.	Case description	Subcases (a) follow	(b) Strict reversal	(c) Weak reversal
1	UTA CH ↓ Forum-write	UTA CH appeared before *any* Forum-write events	A Forum-write event happened during a time *when UTA CH was available* before any UTA CH event	A Forum-write event happened during a time *when UTA CH was unavailable* before any UTA CH event
2	Forum-write ↓ UTA CH	Forum-write appeared before *any* UTA CH events	**UTA CH appeared before any Forum-write events**	-
3	GTA+ CH ↓ Forum-write	GTA+ CH appeared before *any* Forum-write events	A Forum-write event happened during a time *when GTA+ CH was available* before any GTA+ CH event	A Forum-write event happened during a time *when GTA+ CH was unavailable* before any GTA+ CH event
4	Forum-write ↓ GTA+ CH	Forum-write appeared before *any* GTA+ CH events	**GTA+ CH appeared before any Forum-write events**	-
5	UTA CH ↓ GTA+ CH	UTA CH appeared before *any* GTA+ CH events	-	**GTA+ CH appeared before any UTA CH events**
6	GTA+ CH ↓ UTA CH	GTA+ CH appeared before *any* UTA CH events	-	**UTA CH appeared before any GTA+ CH events**

Bolded strict reversal cases (2b and 4b) have no corresponding weak reversal cases since the forum was always accessible. Bolded weak reversal cases (5c and 6c) have no corresponding strict reversal cases since UTA CH and GTA+ CH had exclusively non-overlapping time slots in all classes used in this analysis. Cases 3–6 are not tracked in Python in Fa23-Sp24 due to GTA+ CH data not being collected.

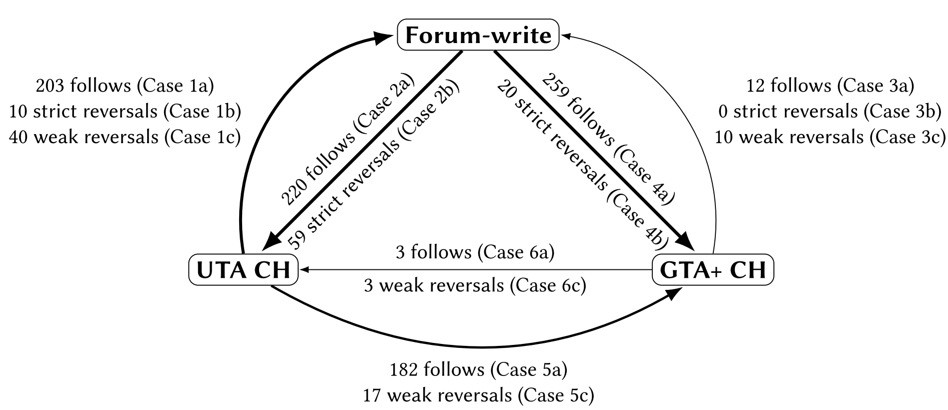


Fig. 2. Number of sequences across all six classes that follow or reverse students' self-reported pairwise orders of resource usage. Each case of pairwise order between clusters is illustrated by an arrow that goes from the resource cluster the student reported using earlier to the cluster that they reported using later. The thicknesses of the arrows are proportional to the number of follows and reversals. Note that a sequence can follow and/or reverse multiple orders at once, so the total numbers of follows and reversals do not sum up to the number of total help-seeking sequences.

Figure 3 plots the percentages of all help-seeking sequences categorized as either follows, strict reversals, or weak reversals for each case *per class*. For example, 51 (17.5%), 14 (4.8%), and 6 (2.1%) of the 292 help-seeking sequences in Python-Sp23 are Case 1 follows, weak reversals, and strict reversals, respectively; these are the three leftmost bars in the top-left subfigure for Python-Sp23. It is important to note that, although we could obtain the proportion of help-seeking sequences that fall into each separate case, the *total* number of follows or reversals should *not* be treated as

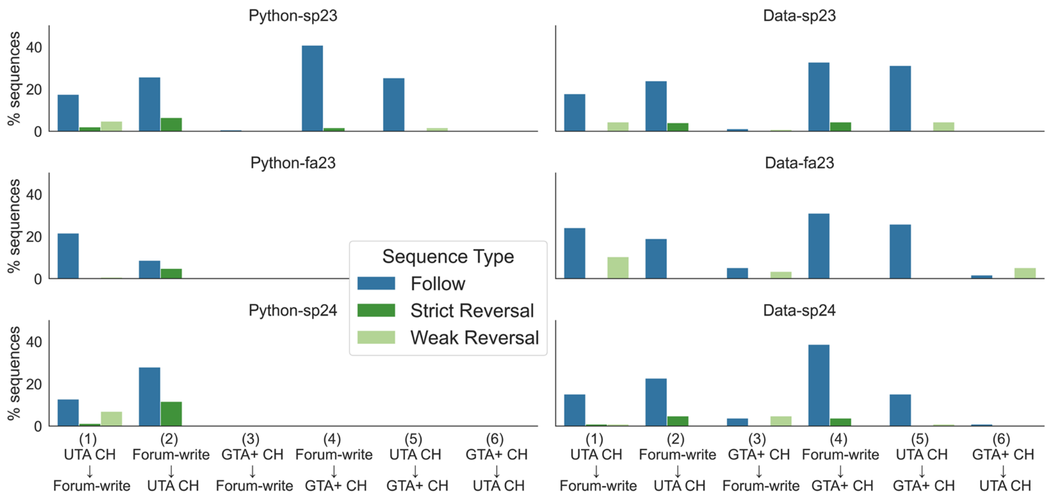


Fig. 3. Bar plot of sequences that follow or reverse students' self-reported pairwise orders of resource usage. Note that all bars show percentages of help-seeking sequences for that class (sequences in which at least one of UTA CH, GTA+ CH, or Forum-write appeared). Cases 3 to 6 were not tracked in Python in Fa23-Sp24. Note that a sequence can follow or reverse one or more cases, and thus the heights for all bars for a class do not sum up to 100.

proportions since a student-assignment-level sequence could lead to multiple follows and reversals at once (e.g., if the student indicated they would use both UTA CH and GTA+ CH strictly before Forum-write, but wrote on the forum first when no CH was available, this would count towards Case 1c *and* Case 3c).

The proportions shown in Figure 3 also reflect the relative magnitude of Dominance between the clusters. About half of the students in a typical class ranked Forum-write strictly earlier than GTA+ CH (see Figure 1(b)), the largest proportion between the three clusters investigated here; in correspondence, Case 4 (Forum-write → GTA+ CH) got the largest number of follows *and* the highest follow-to-reversal ratio in all four classes in which GTA+ CH was tracked.

In all, we found that students' actual order of help resource usage, *when limited to those resources observable in our official records*, mostly matched their self-reported approaches in the Order survey. This suggests that the clusters in Section 5 are adequately indicative of the sequentiality in students' help-seeking behavior.

7 RQ3: Mining Frequent Sequential Rules in Student-Assignment-Level Sequences

We conducted frequent sequential rule mining [18, 30] on the student-assignment-level sequences constructed in Section 4.3.4. We focused on mining frequent *sequential rules* (i.e., given that some set of one or more event(s) happened, another set of one or more event(s) frequently happened subsequently) rather than *sequential patterns* (i.e., some sequences of events occurred frequently) because the latter is prone to favor more common events in imbalanced datasets (e.g., autograder attempts in our context).

7.1 Procedure

To unify the instructional contexts in all classes, we assigned a single code to all kinds of CHs (in-person or online, UTA CH or GTA+ CH) due to the fluctuating instructional context (e.g., GTA+ CH not captured in Python-Fa23 and Python-Sp24, or UTA CH being offered in hybrid modality in some classes but not others; see Table A2). We also treated all *Autograder (all tests passed)* events as *Autograder (progress)* events due to the former being available only in Gradescope-hosted assignments (see Section 4.3.4).

We mined sequential rules of the form $R : \mathbf{X} \rightarrow \mathbf{Y}$ where $\mathbf{X}$ and $\mathbf{Y}$ represent *unordered sets* of one or more event(s). The meaning of such a rule is that if *all events in* $\mathbf{X}$ occur (in any arbitrary order) in a sequence, they will (frequently) be followed by all events in $\mathbf{Y}$, in any arbitrary order, in the same sequence. The *support* of a rule R is the number of sequences in the dataset where R is observed. The *confidence* of R is the support of R divided by the number of sequences in which all events in $\mathbf{X}$ occurred, and thus can be thought of as the empirical probability (observed from the dataset) of R conditioned on the occurrence of $\mathbf{X}$.

Separately for each *course*, we obtained the rules via the following process:

- We ran the TopSeqRules [19] algorithm in the SPMF data mining library [18] on the set of sequences in each class, obtaining the top-1,000 rules (ranked by confidence) with a minimum confidence of at least 0.1 along with their confidence and support values.[9]
- We took the *intersection* of the sets of rules across all classes within the course and keep only the rules with at least 50 total occurrences of the antecedent set (the $\mathbf{X}$ in the above). This was to ensure the final results contained only rules that are (1) meaningful in all classes, as opposed to only in a few of them; and (2) with substantial evidence, as opposed to being spurious ones resulting from extremely small sizes of antecedent occurrences.
- For the remaining rules, we then obtained their *overall* confidence and support values by giving uniform weight to each occurrence of the antecedent set. In other words, we pooled the data from all offerings of each course and obtained the rules' confidence and support values from the pooled datasets.[10] This avoided overweighing results from classes with smaller sample sizes (and thus more prone to discovering outlier rules).
- We then filtered the resulting rules as follows:
  - We discarded all rules with at least one of the antecedent set ($\mathbf{X}$) and the consequent set ($\mathbf{Y}$) having three or more event codes. These rules are harder to interpret and are usually *subcases* of simpler rules.[11]
  - We discarded the rules whose consequent set $\mathbf{Y}$ contained only autograder attempts. This was because autograder attempts make up the vast majority of the event pool and largely just represent normal usage.
  - We discarded the *redundant* rules with identical measurements (confidence and support) and implications. These rules were due to every CH visit produces three events corresponding to the request, the start of the interaction, and the end of the interaction. For a concrete example, *CH request* ⟶ *{CH end, Autograder (progress)}* and *CH start* ⟶ *{CH end, Autograder (progress)}* had identical confidences, supports, and implications, and thus we only kept one of them.
  - For *mostly redundant* rules with similar but not completely identical measurements, we *grouped* the rules together, i.e., treating them as conceptually the same rule; concrete examples can be found in Table 9 as well as the discussions below.
  - Next, we categorized the remaining rules into *trivial* and *nontrivial* ones. Rules that represent conventional single usage of a single resource (e.g., *CH request* ⟶ *CH end*, or *Forum thread*

[9]These threshold values are implemented mainly just for efficiency reasons, as the final list of rules discussed in both courses will contain just around 20 rules, all with confidence values of at least 0.25.
[10]Note that the number of sequences varies from class to class (see Table A3 in Appendix Section A.2) due to fluctuations in number of formative assignments captured in our data (Table A2) as well as enrollment/consent rate (Table A1).
[11]For a conceptual example, $\{a, b, c\} \rightarrow \{d, e\}$ is a subcase of $\{a, b\} \rightarrow \{d, e\}$. Unless the addition of c to the antecedent set drastically changes the confidence of the rule, not much insight is lost by ignoring such a rule.

Table 9. Top-10 Frequent Sequential Rule Groups Mined in Python and Data along with Trivial Rules with At Least the Same Confidence

Rank Python	Rank Data	Rule	Python Supp.	Python Conf.	Data Supp.	Data Conf.
1	1	CH request → {CH end, Autograder (progress)}	1,164	0.934	494	0.905
		CH start → {CH end, Autograder (progress)}	1,162	0.933	493	0.903
2	2	{Autograder (no progress), CH request} → {CH end, Autograder (progress)}	554	0.668	264	0.700
		{Autograder (no progress), CH start} → {CH end, Autograder (progress)}	552	0.666	263	0.698
		{Autograder (no progress), CH request} → {CH start, Autograder (progress)}	505	0.609	258	0.684
3	3	Forum thread initiation → {Forum response, Autograder (progress)}	486	0.599	289	0.634
4	5	CH request → {CH end, Autograder (no progress)}	673	0.540	251	0.460
		CH start → {CH end, Autograder (no progress)}	657	0.527	248	0.454
5	10	CH end → CH request	580	0.465	179	0.328
6	-	CH start → {Autograder (progress), CH request}	561	0.450	170	0.311
		CH end → {Autograder (progress), CH request}	556	0.446	170	0.311
7	6	Forum self-comment → {Forum response, Autograder (progress)}	152	0.433	62	0.425
8	9	{Autograder (progress), CH request} → {CH end, Autograder (no progress)}	484	0.394	180	0.350
		{Autograder (progress), CH start} → {CH end, Autograder (no progress)}	471	0.384	178	0.346
		{Autograder (progress), CH request} → {CH start, Autograder (no progress)}	459	0.374	177	0.344
9	4	{Autograder (no progress), Forum thread initiation} → {Forum response, Autograder (progress)}	211	0.362	142	0.467
10	7	Forum thread initiation → {Forum response, Autograder (no progress)}	293	0.361	162	0.355
-	8	{Autograder (no progress), Forum self-comment} → {Forum response, Autograder (progress)}	73	0.289	35	0.354
Trivial		CH request → CH start → CH end	1,246	1.000	546	1.000
		Forum thread initiation → Forum response	549	0.676	326	0.715
		Forum self-comment → Forum response	167	0.476	72	0.493
		Forum thread initiation → Forum self-comment	339	0.417	129	0.283
		Forum response → Forum self-comment	234	0.402	114	0.306

Rules are ordered by their confidence in Python.

initiation → Forum response) were categorized as trivial and carrying little insight. We discarded all the rules that are *subcases of another trivial rule*; for example, *{Autograder (progress), CH request} → CH end* was a subcase of the trivial rule *CH request → CH end* and was thus discarded. All remaining rules that involved multiple resources or multiple usage of a single resource were considered nontrivial.

— Finally, we ranked all remaining trivial and nontrivial rules separately by their confidence. Note that the first step of this procedure already guaranteed a lower bound of the number of occurrences of the antecedent set (at least 50 occurrences).

7.2 Frequent Rules for Python and Data

Table 9 summarizes the top-10 (in terms of confidence) rule *groups* obtained by the procedure above for both Python and Data as well as all trivial rules with confidence levels in this range (i.e., higher than any of the rules in the top-10 groups) as context. Here on, we refer to groups by their ranking in Python unless otherwise noted.

Overall, the results show that actively seeking help from formal social resources led to some kind of measurable progress more often than not. For example, Group 1 (the highest ranked nontrivial rule group in both courses) indicates that in 90% or more of cases where a student accessed a CH, they made some kind of progress afterward (either during or after the CH interaction, as the consequence set is unordered). This can be contextualized by Group 4, the "opposite outcome" of Group 1 which measures the empirical probability that a student would make an autograder attempt without progress after utilizing the CH, having a substantially lower confidence of 45–54%.

The trend of *desirable outcomes ranking higher than undesirable counterparts* is prevalent in the results here: In addition to the above example for general CH usage, Group 3 vs. Group 10 exhibits the same phenomenon for writing on the forum, and Group 2 vs. Group 8 is a slightly more complicated example for CH usage that indicates it is more likely for a student that attended CH to go from no progress to having progress than the other way around. In general, we found no group of rules with undesirable consequences in either course to be both in the top-10 results and ranked higher than its desirable counterpart.

The other frequent rule groups in the top results indicated that *usage habits may better predict help-seeking behavior than any other tangible evidence of help needs*. More specifically, we hypothesized that we would see high confidence values for rules representing students seeking help in CH after being stuck in the autograder. Instead, we found the rule *CH end* → *CH request* (Group 5), with a total confidence of 0.465 in Python and 0.328 in Data, to be the highest ranked rule with *CH request* in the consequence set in both courses. In other words, the best indicator for students' subsequent CH usage is actually that students have already used CH *within the lifecycle of the same assignment*. This hints that *help-seeking leads to more help-seeking* from the same help resource, a feedback loop that may evolve into a usage habit.

Finally, we observed small (but observable) differences in support values of the rules within the same groups. As an example, the difference between the support values of the two rules in Group 1 indicates that in a few sequences (two sequences in Python, one in Data) where the student utilized CH, some autograder progress happened *between the request and the start of the CH interaction* whereas no further progress happened after the start. In contrast, the rules in Group 4 have a larger difference (16 counts in Python, 3 in Data), validating that students were more likely to be in a "stuck" status between requesting a CH interaction and starting it. Similar discrepancies can be observed in Groups 2, 6, and 8, revealing that some students were attempting the autograders not only *while waiting to interact with instructors/TAs* (as reported by Gao et al. [24]) but also *during their interaction with the instructors/TAs*, likely as part of a debugging process. We discuss this topic more in Section 9 as an interesting future research direction.

The top-3 rules in both courses are completely identical. In fact, 9 of the top-10 rules in Data overlap that in Python; the only rule not present in the Python results (ranked 8th in Data) happens to be the rule with the smallest support in both courses. Of particular note is that we found *no* rules in either course involving both formal social resources (CH and the forum) at the same time, indicating that students did not utilize both resources *in the same sequence* often. Indeed, less than 3% of the student-assignment level sequences in *all* classes involved both resources.

8 Discussions and Implications

At a high level, our results showed the existence of sequentiality in computing students' help resource utilization. There is *sequentiality among help resources*: When explicitly asked about the order of their resource usage (RQ1), students' responses to the Order survey revealed a progression of clusters of resources (Section 5.2) that can be explained by various dimensions in the landscape (Section 5.3). Our audit analyses in Section 6 (RQ2) then showed these orders matched students' behavior in formal resources under availability constraints. The frequent rule mining in Section 7 (RQ3) examined the data in finer granularity and factors in autograder records, revealing *sequentiality within the same help episode*: On the same assignment, students used formal resources in immediate succession, and often went back to the same resource.

Sequentiality is not a new concept. A few existing works [10, 14, 89] have both already reported progressions of help resource usage as thematic findings from student interviews. Broader works on students' learning progress have also reported students in introductory computing classes sometimes wait until their most preferred help resources become available [51] while sometimes resorting to certain help resources when their default ones are not available [52]. Treating autograders as a broad (but sometimes passive) help resource, there have been works reporting how problematic autograder designs cause confusion and drive up CH usage [2, 12, 90].

However, the concepts of *sequential behavior*, *preference*, and *usage frequency* are often conflated, partially due to a lack of emphasis on sequentiality in help-seeking theories. Prior works seemed to implicitly assume that students' sequential help-seeking behavior always follows their order of resource preferences because they investigated students' help resource *preference* via the proxy of their *frequencies* of using all resources [14, 33, 89]. However, Doebling and Kazerouni's interviews [14] already revealed that students use resources with low Formality and high accessibility first despite perceiving them as less useful. At the very least, their results showed that the assumption that students have a universal preference over all help resources is oversimplifying; there might not even be a complete order of preference over all resources to begin with.

With more and more resources entering the landscape, instructors should avoid falling into this pitfall of oversimplifying students' help-seeking. For example, instead of characterizing students into formal and informal resource users as a dichotomy, instructors should acknowledge that students often seek help from external resources due to their relatively higher Availability and Timeliness as well as lower Time Anchor and Space Anchor, and may gravitate towards formal resources more if the formal resources are enhanced along the relevant dimensions.

As formal help resources are usually the only part of the landscape that falls under instructors' control, instructors may also consider *emphasizing the less replaceable values in formal resources.* Course-affiliated resources such as CH allow for high Adaptability, high Socialness, and Media richness interactions, allowing affective benefits [51] and increased sense of belonging [64], especially in introductory settings. Adequate training and explicit instructions on non-technical parts of help-giving can help improve formal resources along these lines.

Finally, students do not necessarily have a clear picture of the help landscape in their minds. Lower-division and especially first-generation students may not have the skills to navigate social help resources or the metacognitive skills to determine when to seek help from what resource [46]. It is likely that these students would benefit from explicit instructions on effectively selecting resource(s) to seek help from, where the help landscape and its dimensions may be useful tools to help deconstruct help-seeking.

9 Limitations

9.1 Scope of Study

Demographics and Identities. The gender, race, and year distributions fluctuated from class to class in our data (Table A1). Although we report all of these for context, the potential influence of demographic variables/identities on students' help-seeking is out of the scope of this article, partially due to space and the lack of racial/ethnic diversity in our data. We would like to note that *this is far from claiming that these variables do not have any effect.* Consistent with past findings [14, 23, 74, 84, 85, 93], women sought help more (and more frequently) in our datasets, but the small number of students in some racial/ethnic groups prevented us from further looking at race, let alone intersectionality.

We also chose not to analyze prior programming experience. Multiple existing works have reported both positive [10, 75] and negative [44, 93] evidence of prior programming experiences influencing students' help-seeking behavior. However, our data does not lend itself well to studying prior programming experiences. Students with ample prior programming experience often skip Python altogether at Duke, resulting in a student body consisting of mostly inexperienced students (and potentially, students with some prior experience but lower self-efficacy) in that course. The interdisciplinary nature of Data makes the concept of prior programming experience less applicable, as that course caters to both students with extensive programming experience but little statistics background and students the other way around. Similarly, a significant portion of students in Python and Data take the course before declaring their majors. We thus also chose not to analyze or even report our students' majors.

Non-Assignment-Based Help-Seeking. In approximating each help episode using a student-assignment-level sequence in RQ2–3, we implicitly assumed most of students' help-seeking happened around the assignments. Although this assumption held in our context,[12] our analyses in RQ2–3 shed little light on students' help-seeking in other contexts, e.g., understanding the class material or preparing for exams. It is quite possible that students' help-seeking behavior in those contexts, which cannot be investigated via our approach, is fundamentally different from their assignment-related help-seeking behavior. An interesting future work direction is to examine whether students utilize different (combinations or orders of) resources when seeking goal-based help (e.g., completing an assignment or identifying a bug in their code) vs. seeking conceptual help.

9.2 Analytic Methods

Order of Resource Usage. Our analyses in Section 5 on students' order of resource usage did not control for students' help-seeking frequencies. This means a student who never actively sought help was weighted the same as a student who utilized all help resources heavily in the aggregation. However, we note that we ran the same analyses for only students who utilized formal help resources at least once, and the results did not differ much from that in Section 5.

On comparing students' actual resource utilization order vs. their self-reported approaches in Section 6, our analyses, which depended on the student-assignment-level sequences, only captured the relative order of events but not the timing details. One potential direction is to explicitly look into the *opportunistic waiting time*: Instead of merely categorizing reversals of students' own order of resource usage by whether the earlier/preferred resource was available, one may weight this by *the amount of time students would have needed to wait until the next availability of the earlier/preferred resource*,[13] although that approach might then suffer from not capturing students' own schedules and availabilities.

Mining Students' Help-Seeking Logs. The insights from our frequent rule mining in Section 7 are fundamentally limited by the kind of events we tracked in our data. More insights might emerge had we been able to track reading the forum or utilizing any informal resources at the assignment level. To unify our fluctuating instructional contexts throughout different offerings of the same course, we chose not to distinguish between different CH modalities (in-person, online, or even hybrid; see Table A2), between short and long CH interactions, and between different types of autograders.

Due to the nature of constructing the sequences via only the chronological order of events but not the exact timestamps, our mining could not distinguish between transient sequential event patterns (e.g., students submitting to the autograder several minutes after completing a CH interaction) vs. long-term sequential event patterns (e.g., when this gap was 2 days long, during which other informal forms of help might have happened). Explicitly analyzing the time elapsed between different types of events like the works by Gao et al. around CH [21, 24] might help shed light on several different questions: (1) How long may it take a student to *feel stuck* (e.g., by not getting

[12]As mentioned in Section 4.3.1, 86.4% of valid Python and 87.3% of Data CH interactions were tied to a specific assignment. Also see Table A3.

[13]Note that this concept is fundamentally different from the amount of time students are willing to wait in line to get helped [22, 24].

any measurable progress in the autograder) before they decide to seek help? (2) *Assuming students would utilize help resources sequentially according to some order*, how long does it take a student to move on from trying to access (or waiting to access) a certain help resource (cluster) and resort to the next resource (cluster), and how is this affected by resource availability? (3) *How long does a "learning effect" from help-seeking last*? (4) How much, how, and why do students attempt the autograder during waiting for CH interaction [24] or even during the CH interaction itself? Each of the questions above is worth its own fine-grained analysis and we leave all for future work.

9.3 Data Collection

Instructional Context. Enrollment and consent rates fluctuated among classes in our data (Table A1). Women had a higher consent rate than men in most of the datasets, potentially introducing a selection bias. While the contents in Python and Data (courses used for analyses in RQ2–3) were generally stable, some details, such as the relative order of topics, changed over time. The course-level analyses we conducted and presented assumed these factors were negligible. However, we repeated all course-level analyses at the class level and did not find any anomalies in which one class's data showed drastically different behavior.

The Order survey. Our survey question on the order of help resource usage was administered at slightly different points during the semester in each class. Therefore, we could not confidently assume their responses captured their order of resource usage *for any specific help episode* or *overall for the entire semester*, but nevertheless still compared their responses with their actual usage logs in Section 6. Furthermore, although our survey question was phrased explicitly to take the consideration of the availability of resources away, it is unclear whether the students followed this instruction. As seen in the analyses in Section 5, there is substantial evidence that the availability of help resources either directly or indirectly affects students' help-seeking approaches. Finally, the lack of presence of the Order survey in earlier semesters made it infeasible to conduct the auditing analyses in Section 6 in those semesters.

Data from Help Platforms. Each data source has its own limitations. For CH, 10.4% and 8.3% of consenting MyDigitalHand interactions respectively in Python and Data had invalid timestamps (see Table 5), and it is not known how many interactions were not recorded due to human error/neglect. Instructor/GTA CH was captured in only 4/8 of the 11 classes, respectively (Table A2), rendering Cases 3–6 not auditable in some classes (Section 6.2).

For the forum, we only collected students' "writing" actions but not their "reading" actions (i.e., viewing others' posts),[14] which is nonetheless an essential part of their help-seeking behavior [8, 75, 80, 94]. The forum datasets also did not include threads that were *deleted*; anecdotally, we observed students self-removing their forum threads after getting satisfactory responses (or after turning to CH and receiving help there).

As students worked on all assignments locally instead of in a controlled IDE, our autograder datasets only captured the submission actions instead of all engagements with the assignments. Although students were explicitly taught to submit their work incrementally, some embraced the style of finishing all work locally and submitting at once. Although not included in this article, the distributions of number of autograder attempts per student in each class have very wide ranges. Especially for longer assignments administered on Gradescope, for which sometimes the autograders took several minutes to finish a single run through the entire assignment regardless of

[14]Ed Discussion records partial reading behavior, e.g., the number of unique posts a student read over the semester, but not at the *contribution- or thread-level*, so we could not infer the timestamp a student read a specific contribution (or whether that happened at all).

the scope of each update, we have anecdotal evidence that students did not make an autograder attempt until they thought they had made *enough* progress. Therefore, an autograder attempt might have a different meaning to students with different submission habits.

Finally, not all events were annotated with assignment information. For a small amount (about 2%) of unannotated forum contributions, the available information (see Table A4 in Appendix Section A.2) did not successfully point to an assignment; we resorted to comparing their timestamps with assignment deadlines. As such, in some rare cases, students with personal extensions for their assignments or were retroactively asking a question about a past-due assignment might have their forum contributions misclassified, introducing errors into the sequences.

10 Conclusions

We studied post-secondary computing students' academic help-seeking approach and behavior via their self-reported order of resource usage and their records in course-affiliated help resources. Students' orders of resource usage form a progression of clusters (Section 5) that confirms and expands phenomena reported by existing literature, and students' observable actions mostly match their approaches (Section 6). We mined frequent rules in student-assignment-level help-seeking event sequences (Section 7), in which we saw a prominent phenomenon of usage habit better predicting future help-seeking behavior.

Our work is only a first step in attempting to analyze computing students' help-seeking approach and behavior with an awareness of resource sequentiality. We demonstrated that rich insights can be found with this approach even within our limited data scope, while many interesting RQs remain in this direction (Sections 8 and 9). Understanding the complex interconnection between help resources and how students use them would benefit large-scale resource-rich computing course contexts in their resource allocation/management as well as their help ecosystem design.

Acknowledgments

We thank the numerous instructors, teaching associates, and departmental staff for their efforts in help collecting and transferring the data. We thank Melissa Chen, Yinmiao Li, and the anonymous reviewers for their helpful and constructive feedback to the article at various stages.

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. DOI: 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. DOI: https://doi.org/10.1145/2724660.2724679

[3] Alejandro Bogarín, Rebeca Cerezo, and Cristóbal Romero. 2018. A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery* 8, 1 (2018). DOI: https://doi.org/10.1002/widm.1230

[4] Alejandro Bogarín, Cristóbal Romero, Rebeca Cerezo, and Miguel Sánchez-Santillán. 2014. Clustering for improving educational process mining. In *ACM LAK*, 11–15. DOI: https://doi.org/10.1145/2567574.2567604

[5] 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. DOI: https://doi.org/10.1145/3478431.3499385

[6] Jennifer Campbell and Michelle Craig. 2018. Drop-in help centres: An alternative to office hours. In *ACM WCCCE*, 9:1–9:6. DOI: https://doi.org/10.1145/3209635.3209642

[7] Rebeca Cerezo, Alejandro Bogarín, Maria Esteban, and Cristóbal Romero. 2020. Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education* 32, 1 (2020), 74–88.

[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. DOI: https://doi.org/10.1145/3576050.3576106

[9] codePost. 2025. Retrieved from https://codepost.io/

[10] Anael Kuperwajs Cohen, Alannah Oleson, and Amy J. Ko. 2024. Factors influencing the social help-seeking behavior of introductory programming students in a competitive university environment. *ACM Transactions on Computing Education* 24, 1 (2024), 11:1–11:27. DOI: https://doi.org/10.1145/3639059

[11] 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. *Computing Education* 162 (2021), 104094. DOI: https://doi.org/10.1016/j.compedu.2020.104094

[12] Andrew Deorio and Christina Keefer. 2022. When is automated feedback a barrier to timely feedback? In *ASEE*. Retrieved from https://peer.asee.org/40405

[13] Ed Discussion. 2025. Retrieved from https://edstem.org/

[14] Augie Doebling and Ayaan M. Kazerouni. 2021. Patterns of academic help-seeking in undergraduate computing students. In *ACM Koli Calling*, 13:1–13:10. DOI: https://doi.org/10.1145/3488042.3488052

[15] Radim Dolak. 2019. Using process mining techniques to discover student's activities, navigation paths, and behavior in LMS Moodle. In *ICITL*, 129–138. DOI: https://doi.org/10.1007/978-3-030-35343-8_14

[16] 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. DOI: https://doi.org/10.1145/3369255.3369258

[17] 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. DOI: https://doi.org/10.1037/edu0000725

[18] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Antonio Gomariz, Ted Gueniche, Azadeh Soltani, Zhihong Deng, and Hoang Thanh Lam. 2016. The SPMF open-source data mining library version 2. In *ECML PKDD*, 36–40. DOI: https://doi.org/10.1007/978-3-319-46131-1_8

[19] Philippe Fournier-Viger and Vincent S. Tseng. 2011. Mining top-K sequential rules. In *ADMA*, 180–194. DOI: https://doi.org/10.1007/978-3-642-25856-5_14

[20] 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. DOI: https://doi.org/10.5281/zenodo.6853172

[21] Zhikai Gao, Bradley Erickson, Yiqiao Xu, Collin Lynch, Sarah Heckman, and Tiffany Barnes. 2022. You asked, now what? Modeling students' help-seeking and coding actions from request to resolution. *Journal of Educational Data Mining* 14, 3 (2022), 109–131. DOI: https://doi.org/10.5281/zenodo.7304908

[22] Zhikai Gao, Adam M. Gaweda, Collin F. Lynch, Sarah Heckman, Damilola Babalola, and Gabriel Silva de Oliveira. 2024. Using survival analysis to model students' patience in online office hour queues. In *ACM SIGCSE TS*, 1646–1647. DOI: https://doi.org/10.1145/3626253.3635517

[23] 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. DOI: https://doi.org/10.1145/3478431.3499334

[24] Zhikai Gao, Collin Lynch, and Sarah Heckman. 2023. Too long to wait and not much to do: Modeling student behaviors while waiting for help in online office hours. In *CSEDM Workshop*. DOI: https://doi.org/10.5281/zenodo.7761609

[25] Zhikai Gao, Collin Lynch, Sarah Heckman, and Tiffany Barnes. 2021. Automatically classifying student help requests: A multi-year analysis. In *EDM*, 81–92. Retrieved from https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_153.pdf

[26] 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. *Journal of Computing in Higher Education* 33, 1 (2021), 1–18. DOI: https://doi.org/10.1007/s12528-020-09252-0

[27] GitHub Copilot. 2025. Retrieved from https://github.com/features/copilot/

[28] Gradescope. 2025. Retrieved from https://www.gradescope.com/

[29] Georgiana Haldeman, Monica Babes-Vroman, Andrew Tjang, and Thu D. Nguyen. 2021. CSF: Formative feedback in autograding. *ACM Transactions on Computing Education* 21, 3 (2021), 21:1–21:30. DOI: https://doi.org/10.1145/3445983

[30] Jiawei Han, Micheline Kamber, and Jian Pei. 2011. *Data Mining: Concepts and Techniques* (3rd. ed.). Retrieved from http://hanj.cs.illinois.edu/bk3/

[31] 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. *Computer Science Education* 32, 1 (2022), 105–127. DOI: https://doi.org/10.1080/08993408.2020.1860408

[32] 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. DOI: https://doi.org/10.1145/3568813.3600139

[33] 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. DOI: https://doi.org/10.1145/3636243.3636248

[34] Iris K. Howley and Carolyn P. Rosé. 2018. Empirical evidence for evaluation anxiety and expectancy-value theory for help sources. In *ICLS*. Retrieved from https://repository.isls.org//handle/1/891

[35] Andrew Jiang and Bogdan Simion. 2022. Help supports during online delivery: Student perception and lessons learnt from an online CS2. In *ACM SIGCSE TS*, 105–111. DOI: https://doi.org/10.1145/3478431.3499369

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

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

[38] 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. Retrieved from https://www.infoagepub.com/products/Applications-of-Self-Regulated-Learning-across-Diverse-Disciplines

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

[40] 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. DOI: https://doi.org/10.1037/0022-0663.83.2.221

[41] Shamya Karumbaiah, Jaclyn Ocumpaugh, and Ryan Shaun Baker. 2022. Context matters: Differing implications of motivation and help-seeking in educational technology. *International Journal of Artificial Intelligence in Education* 32, 3 (2022), 685–724. DOI: https://doi.org/10.1007/s40593-021-00272-0

[42] 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. DOI: https://doi.org/10.1145/3544548.3580919

[43] Hieke Keuning, Johan Jeuring, and Bastiaan Heeren. 2019. A systematic literature review of automated feedback generation for programming exercises. *ACM Transactions on Computing Education* 19, 1 (2019), 3:1–3:43. DOI: https://doi.org/10.1145/3231711

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

[45] Shao-Heng Ko and Kristin Stephens-Martinez. 2024. The trees in the forest: Characterizing computing students' individual help-seeking approaches. In *ACM ICER*, 343–358. DOI: https://doi.org/10.1145/3632620.3671099

[46] Shao-Heng Ko, Kristin Stephens-Martinez, Matthew Zahn, Yesenia Velasco, Lina Battestilli, and Sarah Heckman. 2025. Student perceptions of the help resource landscape. In *ACM SIGCSE TS*, 596–602. DOI: https://doi.org/10.1145/3641554.3701851

[47] 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. DOI: https://doi.org/10.1145/3502718.3524786

[48] 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. DOI: https://doi.org/10.1145/3478432.3499099

[49] 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. DOI: https://doi.org/10.1145/3568813.3600138

[50] 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. DOI: https://doi.org/10.1145/3478431.3499372

[51] Yinmiao Li, Melissa Chen, Ayse Hunt, Haoqi Zhang, and Eleanor O'Rourke. 2024. Exploring the interplay of metacognition, affect, and behaviors in an introductory computer science course for non-majors. In *ACM ICER*, 27–41. DOI: https://doi.org/10.1145/3632620.3671119

[52] Yinmiao Li, Haoqi Zhang, and Eleanor O'Rourke. 2024. The undervalued disciplinary and emotional support provided by teaching assistants in introductory computer science courses. In *ICLS*, 1498–1501. DOI: https://doi.org/10.22318/icls2024.614567

[53] 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 Transactions on Computing Education* 22, 4 (2022), 39:1–39:31. DOI: https://doi.org/10.1145/3487050

[54] Jorge Javier Maldonado Mahauad, Mar Pérez-Sanagustín, René F. Kizilcec, Nicolás Morales, and Jorge Munoz-Gama. 2018. Mining theory-based patterns from Big Data: Identifying self-regulated learning strategies in massive open online courses. *Computers in Human Behavior* 80 (2018), 179–196. DOI: https://doi.org/10.1016/j.chb.2017.11.011

[55] 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. Retrieved from https://www.infoagepub.com/products/Advances-in-Help-Seeking-Research-and-Applications
[56] Amogh Mannekote, Mehmet Celepkolu, Aisha Chung Galdo, Kristy Elizabeth Boyer, Maya Israel, Sarah Heckman, and Kristin Stephens-Martinez. 2022. Don't just paste your Stacktrace: Shaping discussion forums in introductory CS courses. In *ACM SIGCSE TS*, 1164–1164. DOI: https://doi.org/10.1145/3478432.3499110
[57] Julia M. Markel and Philip J. Guo. 2021. Inside the mind of a CS undergraduate TA: A firsthand account of undergraduate peer tutoring in computer labs. In *ACM SIGCSE TS*, 502–508. DOI: https://doi.org/10.1145/3408877.3432533
[58] 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. DOI: https://doi.org/10.1145/3372782.3406264
[59] Diba Mirza, Phillip T. Conrad, Christian Lloyd, Ziad Matni, and Arthur Gatin. 2019. Undergraduate teaching assistants in computer science: A systematic literature review. In *ACM ICER*, 31–40. DOI: https://doi.org/10.1145/3291279.3339422
[60] MyDigitalHand. 2025. Retrieved from https://beta.mydigitalhand.org/
[61] Sadia Nawaz, Namrata Srivastava, Ji Hyun Yu, Anam Ahmad Khan, Gregor E. Kennedy, James Bailey, and Ryan Shaun Baker. 2022. How difficult is the task for you? Modelling and analysis of students' task difficulty sequences in a simulation-based POE environment. *International Journal of Artificial Intelligence in Education* 32, 2 (2022), 233–262. DOI: https://doi.org/10.1007/s40593-021-00242-6
[62] OpenAI Codex. 2025. Retrieved from https://openai.com/blog/openai-codex
[63] Aadarsh Padiyath, Xinying Hou, Amy Pang, Diego Viramontes Vargas, Xingjian Gu, Tamara Nelson-Fromm, Zihan Wu, Mark Guzdial, and Barbara Ericson. 2024. Insights from social shaping theory: The appropriation of large language models in an undergraduate programming course. In *ACM ICER*, 114–130. DOI: https://doi.org/10.1145/3632620.3671098
[64] Leah Perlmutter, Jean Salac, and Amy J. Ko. 2023. "A field where you will be accepted": Belonging in student and TA interactions in post-secondary CS education. In *ACM ICER*, 356–370. DOI: https://doi.org/10.1145/3568813.3600128
[65] Piazza. 2025. Retrieved from https://piazza.com/
[66] James Prather, Lauren E. Margulieux, Jacqueline Whalley, Paul Denny, Brent N. Reeves, Brett A. Becker, Paramvir Singh, Garrett B. 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.
[67] James Prather, Brent N. Reeves, Paul Denny, Brett A. Becker, Juho Leinonen, Andrew Luxton-Reilly, Garrett B. Powell, James Finnie-Ansley, and Eddie Antonio Santos. 2024. "It's weird that it knows what I want": Usability and interactions with copilot for novice programmers. *ACM Transactions on Computer-Human Interaction* 31, 1 (2024), 4:1–4:31.
[68] James Prather, Brent N. Reeves, Juho Leinonen, Stephen MacNeil, Arisoa S. Randrianasolo, Brett A. Becker, Bailey Kimmel, Jared Wright, and Ben Briggs. 2024. The widening gap: The benefits and harms of generative AI for novice programmers. In *ACM ICER*, 469–486. DOI: https://doi.org/10.1145/3632620.3671116
[69] 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. DOI: https://doi.org/10.1145/3105726.3106179
[70] 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. DOI: https://doi.org/10.1186/s41239-018-0100-7
[71] Yanyan Ren, Shriram Krishnamurthi, and Kathi Fisler. 2019. What help do students seek in TA office hours? In *ACM ICER*, 41–49. DOI: https://doi.org/10.1145/3291279.3339418
[72] Saman Zehra Rizvi, Bart Rienties, Jekaterina Rogaten, and René F. Kizilcec. 2020. Investigating variation in learning processes in a FutureLearn MOOC. *Journal of Computing in Higher Education* 32, 1 (2020), 162–181. DOI: https://doi.org/10.1007/s12528-019-09231-0
[73] 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. DOI: https://doi.org/10.58459/rptel.2023.18017
[74] 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. DOI: https://doi.org/10.1145/3568813.3600121
[75] Naaz Sibia, Angela M. Zavaleta Bernuy, Tiana V. Simovic, Chloe Huang, Yinyue Tan, Eunchae Seong, Carolina Nobre, Daniel Zingaro, Michael Liut, and Andrew Petersen. 2024. Exploring the effects of grouping by programming experience in Q&A forums. In *ACM ICER*, 206–221. DOI: https://doi.org/10.1145/3632620.3671107
[76] Naaz Sibia, Angela M. Zavaleta Bernuy, Joseph Jay Williams, Michael Liut, and Andrew K. Petersen. 2023. Student usage of Q&A forums: Signs of discomfort? In *ACM ITiCSE*, 33–39. DOI: https://doi.org/10.1145/3587102.3588842

[77] James Skripchuk, John Thomas Bacher, and Thomas W. Price. 2024. An investigation of the drivers of novice programmers' intentions to use web search and GenAI. In *ACM ICER*, 487–501. DOI: https://doi.org/10.1145/3632620.3671112
[78] James Skripchuk, Neil Bennett, Jeffrey Zhang, Eric Li, and Thomas W. Price. 2023. Analysis of novices' web-based help-seeking behavior while programming. In *ACM SIGCSE TS*, 945–951. DOI: https://doi.org/10.1145/3545945.3569852
[79] 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. DOI: https://doi.org/10.1145/3017680.3017800
[80] 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. DOI: https://doi.org/10.1186/s41239-020-00200-7
[81] Stack Overflow. 2025. Retrieved from https://stackoverflow.com/
[82] 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. Retrieved from https://www.ijee.ie/1atestissues/Vol35-6A/16_ijee3834.pdf
[83] Sadhana Suryadevara and Kristin Stephens-Martinez. 2022. UPIC a problem-solving framework: Understand, plan, implement, and correctness/debugging. In *ACM ICER*, 50–51. DOI: https://doi.org/10.1145/3501709.3544286
[84] 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. DOI: https://doi.org/10.1145/3408877.3432395
[85] Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik. 2014. Gender, representation and online participation: A quantitative study. *Interacting with Computers* 26, 5 (2014), 488–511. DOI: https://doi.org/10.1093/iwc/iwt047
[86] 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. DOI: https://doi.org/10.1145/3017680.3017745
[87] Kevin Wang and Ramon Lawrence. 2024. HelpMe: Student help seeking using office hours and email. In *ACM SIGCSE TS*, 1388–1394. DOI: https://doi.org/10.1145/3626252.3630867
[88] Matthew West, Geoffrey L. Herman, and Craig Zilles. 2015. PrairieLearn: Mastery-based online problem solving with adaptive scoring and recommendations driven by machine learning. In *ASEE*, 1238. DOI: https://doi.org/10.18260/p.24575
[89] 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. DOI: https://doi.org/10.1080/22054952.2018.1525889
[90] 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. DOI: https://doi.org/10.1145/3488042.3488072
[91] Stephanie Yang, Hanzhang Zhao, Yudian Xu, Karen Brennan, and Bertrand Schneider. 2024. Debugging with an AI tutor: Investigating novice help-seeking behaviors and perceived learning. In *ACM ICER*, 84–94. DOI: https://doi.org/10.1145/3632620.3671092
[92] H. Peyton Young. 1974. An axiomatization of Borda's rule. *Journal of Economic Theory* 9, 1 (1974), 43–52. DOI: https://doi.org/10.1016/0022-0531(74)90073-8
[93] Matthew Zahn, Lina Battestilli, and Sarah Heckman. 2022. Academic help seeking patterns in introductory computer science courses. In *ASEE*. Retrieved from https://peer.asee.org/41526
[94] Matthew Zahn and Sarah Heckman. 2023. Observations on student help-seeking behaviors in introductory computer science courses. In *ACM SIGCSE TS*, 1380. DOI: https://doi.org/10.1145/3545947.3576325
[95] 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. DOI: https://doi.org/10.1007/978-1-4612-3618-4

Appendices

A Additional Contexts

A.1 Participant Demographics

Table A1 reports the number of (consenting) students and their demographic distribution (if collected). Due to privacy requirements, we replaced all cells with a count less than 5 with an asterisk. Note that students can indicate their preference not to answer any question, so the cells with asterisks cannot be inferred from the total and other cells. The consent process also introduces a selection bias as women consented more than men.

Table A1. Student Demographics

Inst.	Course-semester		# Total students	Consenting (consent rate)	Gender			Race					Year			
					Men	Women	Nonbinary	White	Asian	Black	2+	Other	1	2	3	4+
NCSU	MATLAB	Sp24	294	153 (52.0%)	94	42	*	97	14	*	8	*	78	40	13	*
Duke	PYTHON	Fa21	241	177 (73.4%)	64	94	*	80	48	15	13	*	87	48	14	12
		Sp22	221	152 (68.8%)	48	76	*	63	38	*	15	*	81	27	10	6
		Fa22	262	183 (69.9%)	-	-	-	-	-	-	-	-	106	32	23	14
		Sp23	218	163 (74.8%)	63	90	*	90	35	11	14	*	101	25	16	16
		Fa23	262	192 (73.3%)	65	103	5	67	67	17	12	*	95	48	14	16
		Sp24	171	77 (45.0%)	19	53	*	36	19	8	6	*	43	15	9	5
	DSA	Fa23	390	272 (69.7%)	167	97	5	100	131	14	13	*	157	93	15	6
		Sp24	300	177 (59.0%)	57	72	*	55	59	*	9	*	101	21	7	5
	DATA	Sp22	209	145 (69.4%)	84	55	*	59	63	8	6	*	10	56	46	27
		Fa22	208	152 (73.1%)	83	65	*	41	83	8	13	*	*	34	54	58
		Sp23	234	160 (68.4%)	92	61	*	63	67	7	14	5	21	68	50	20
		Fa23	82	66 (80.5%)	33	32	*	22	33	*	6	*	*	19	18	29
		Sp24	160	115 (71.9%)	57	57	*	43	46	13	9	*	13	53	32	17
	DM	Fa23	120	94 (78.3%)	49	42	*	26	45	5	9	*	5	42	37	10
		Sp24	138	71 (51.4%)	34	32	*	17	43	*	*	*	21	38	8	*
	DB	Fa23	330	183 (55.5%)	104	74	*	61	97	6	7	*	*	34	68	
	ALGO	Fa23	160	123 (76.9%)	83	34	*	45	59	8	5	*	*	9	36	73
		Sp24	316	220 (69.6%)	138	78	*	89	95	11	17	*	*	49	120	49

The consent rate is the percentage of consenting students. For all demographic subcategories, the number in the cell is the number of consenting students in that category. Black includes African American, and 2+ stands for multiracial. Not all data was collected in all classes, and not all students answered all questions. Values lower than 5 are replaced with an asterisk. NCSU treats Latinx as a race (classified under Other) while Duke separates the concept from race.

A.2 Course-Affiliated Help Resource Contexts

Tables A2 and A3 summarize the contexts of the internal help resources in PYTHON and DATA used in the analyses in RQ2 and RQ3. Table A4 lists the attributes in the forum data.

We had to infer the assignment associated with forum threads only in PYTHON-Fa21 and DATA-Sp22/Fa22, as the other classes had the students indicate the assignments in `subcategory`. Given a thread with `category=assignments` in these classes, we first examine if `title` contained a keyword followed by a number (e.g., `assignment2`, `assign03`, or `hw4`); if so, all contributions in the same thread are labeled with that assignment. We then repeat the same process on `text` if `title` does not contain such patterns. Finally, for the remaining (roughly 2%) of the data that is still unlabeled, we assigned each record to the ongoing assignment according to `created_at` (i.e., the timestamp of the thread initiation).

Table A2. Summary of Course-Affiliated Help Resources in Classes in Which Behavioral Data Was Collected

Course-	semester	CH modality			# of CH per week			Class forum	Autograding	
		In-person	Hybrid	Online	Instructor(s)	GTAs	UTAs	Platform	# Homeworks	# Projects
Python	Fa21	◎	◎		-	4	15	Ed Discussion	8	6
	Sp22	◎	◎		-	4	15	Ed Discussion	8	6
	Fa22	◎		◎	-	4	20	Ed Discussion	7	6
	Sp23	◎		◎	-	4	20	Ed Discussion	7	6
	Fa23	◎	◎	◎	-	-	25	Ed Discussion	6	6
	Sp24	◎	◎	◎	-	-	20	Ed Discussion	6	6
Data	Sp22	◎		◎	-	-	20	Ed Discussion	9	-
	Fa22	◎		◎	2	4	20	Ed Discussion	9	-
	Sp23	◎		◎	2	6	15	Ed Discussion	9	-
	Fa23	◎		◎	1.5	4	10.5	Ed Discussion	9	-
	Sp24	◎	◎	◎	1.5	4	15	Ed Discussion	9	-

An ◎ represents that some CH were offered in that modality in the class. Hybrid CH means there were TAs helping students in person and online *simultaneously* during a shift. The list of hours only captures the number of hours during a typical week when some form of CH was available, but not how many staff members were available at a given shift. CH that did not collect data (mostly instructor ones) are not listed.

Table A3. Summary of Behavioral Datasets by Class

Course	Semester	CH (# of interactions)				Forums (# of contributions)						Autograders (# of attempts)		Sequences constructed
								Help-seeking		Assignments				
		Raw	Consenting	Valid	Assign-ments	Raw	Consenting	Student	Staff	Student	Staff	Homework	Project	
Python	Fa21	1,006	740 (73.6%)	653	576	2,203	1,779 (80.7%)	600	396	369	172	27,329	6,774	2,547
	Sp22	574	458 (79.8%)	406	342	1,699	1,160 (68.3%)	519	399	262	115	18,878	5,493	2,035
	Fa22	978	697 (71.3%)	608	515	2,934	2,253 (76.8%)	1,095	862	442	246	24,882	7,870	2,340
	Sp23	477	363 (76.1%)	283	261	1,757	1,569 (89.3%)	736	653	355	286	21,215	6,772	2,092
	Fa23	990	906 (91.5%)	805	704	2,413	1,903 (78.9%)	848	731	353	232	20,272	9,106	1,114
	Sp24	263	243 (92.4%)	210	166	1,340	668 (49.9%)	257	319	131	63	8,317	2,797	422
	Total	4,288	3,407 (79.5%)	2,965	2,564	12,346	9,332 (75.6%)	4,055	3,360	1,912	1,114	120,893	38,812	10,550
Data	Sp22	354	225 (63.6%)	202	188	1,607	1,282 (79.8%)	446	375	243	167	10,489	-	1,297
	Fa22	514	326 (63.4%)	293	262	1,733	1,445 (83.4%)	434	421	217	189	9,685	-	1,356
	Sp23	535	364 (68.0%)	333	291	1,982	1,527 (77.0%)	448	493	212	163	9,937	-	1,429
	Fa23	107	94 (87.8%)	74	60	1,129	899 (79.6%)	153	232	46	41	3,511	-	585
	Sp24	165	149 (90.3%)	116	88	1,449	1,046 (72.2%)	301	361	92	75	5,455	-	893
	Total	1,675	1,158 (69.1%)	1,018	889	7,900	6199 (78.5%)	1,782	1,882	760	635	39,077	-	5,560

Numbers in parentheses are proportions over raw data counts. Failed autograder runs (e.g., submissions with wrong files or server technical issues) are omitted from the counts. We do not report raw data counts for autograder datasets as the raw datasets contain many entries from course staff for testing. Forum contributions do not include edits to existing content.

Table A4. Summary of Attributes of Forum Contributions Used

Name	Explanation
`category/subcategory`	Self-annotation of what component of class the thread is about
`type`	Type of contribution (question, announcement, answer, comment)
`created_at`	Timestamp of contribution
`user.role`	Role of the user (student/staff)
`initiated_by`	Role of the user who initialized the thread
`title`	Title of thread
`text`	Text of contribution

B Per-Class Results and Visualizations for RQ1

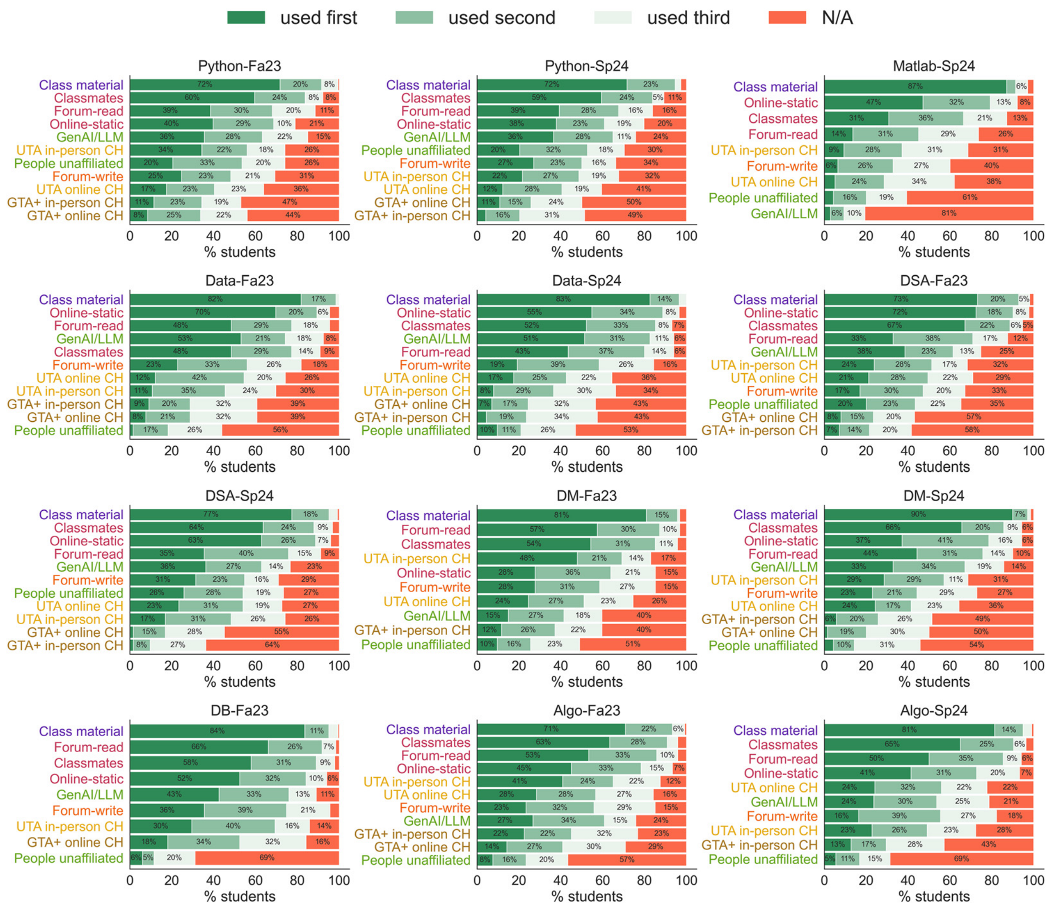


Fig. B1. Students' responses to the Order survey in all classes.

Received 22 November 2023; revised 31 October 2024; accepted 23 January 2025