Relationship Between Help-seeking Behavior of CS Undergraduate Students and Academic Performance

Eunoh Cho

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science and Application

Clifford A Shaffer, Chair
Margaret O Ellis
Sara Hooshangi

May 10,2022
Blacksburg, Virginia

Keywords: Help Seeking Behaviour, CS Student Self Management, Student Skill, Data Anlaysis, Self Management

Relationship Between Help-seeking Behavior of CS Undergraduate Students and Academic Performance

Eunoh Cho

ABSTRACT

Computer Science students need to understand the mechanism of programming systems that involve computation, automation, and information. Computer scientists need to know how to design and analyze a problem and solve it with an algorithm. We study students' behaviors in CS education to find out patterns of those who need help. Several behaviors are examined: Time Management, Incremental development, Self-checking, Persistence, and Planning. Help-seeking, when done correctly, is known as a good strategy related to self-regulated learning. This behavior includes online searching, coming to office hours for help from instructional staff, and asking instructors and peers publicly on online forums. Some of these sources of help can be tracked more easily than others. We present efforts to collect and analyze data related to the help-seeking behavior of students in a second-semester programming course. The goal of this work is to establish mechanisms that will permit us to collect sufficient data from a variety of sources so that we can determine what help-seeking behavior patterns are associated with successful course outcomes.

Our current data collection efforts are tied in part to the effects of the COVID-19 pandemic, which caused courses to be taught online during our data collection period that normally would be taught face-to-face. Data includes logs of viewing or posting questions to the online forum system Piazza, office hour visit logs, Zoom logs, and grades from the Canvas LMS. We present initial analysis such as comparing course grades with the number of times students received help from instructional staff both in office hours and online forum Piazza.

Relationship Between Help-seeking Behavior of CS Undergraduate Students and Academic Performance

Eunoh Cho

GENERAL AUDIENCE ABSTRACT

Computer Science students need to understand the mechanism of programming systems that involve computation, automation, and information. Computer scientists need to know how to design and analyze a problem and solve it with an algorithm. We study students' behaviors in CS education to find out patterns of those who need help. Several behaviors are examined: Time Management, Incremental development, Self-checking, Persistence, and Planning. Help-seeking, when done correctly, is known as a good strategy related to self-regulated learning. This behavior includes online searching, coming to office hours for help from instructional staff, and asking instructors and peers publicly on online forums. Some of these sources of help can be tracked more easily than others. We present efforts to collect and analyze data related to the help-seeking behavior of students in a second-semester programming course. The goal of this work is to establish mechanisms that will permit us to collect sufficient data from a variety of sources so that we can determine what help-seeking behavior patterns are associated with successful course outcomes.

Our current data collection efforts are tied in part to the effects of the COVID-19 pandemic, which caused courses to be taught online during our data collection period that normally would be taught face-to-face. Data includes logs of viewing or posting questions to the online forum system Piazza, office hour visit logs, Zoom logs, and grades from the Canvas LMS. We present initial analysis such as comparing course grades with the number of times students received help from instructional staff both in office hours and online forum Piazza.

Dedicated to Virginia Tech.

Acknowledgments

This project would be impossible without support from others. First of all, Dr.Shaffer has always supported my idea and guided my master's program, including this thesis. Without Dr.Shaffer, this project would have been impossible. Thank you so much, Dr.Shaffer, for making this possible. Thanks to Professor Ellis, who has supported the research and has given me feedback every meeting and proposed new ideas for another guidance. Also, thanks to Dr.Hooshangi, who is my last committee member, for providing feedback on my thesis so that I could avoid mistakes. Other thanks to Molly and Patrick for helping me out with this research. Without you guys, I would also have been in trouble. Thank you to all students who took CS 2114 and agreed to participate in this research. Thank you, Virginia Tech Writing Center, for helping me with my writing skills. My writing skills have improved to another level by working with Writing Center. Thank you, my friends, who have been there for me when good and not. Lastly, thank you to my family, who has supported me until now. Thank you so much, everyone, for supporting me in finishing my master's degree.

Contents

List of Figures

List of Tables

List of Abbreviations

CS Computer Science

HP High Performing

LP Low Performing

Chapter 1

Introduction

While understanding the course content is important, good academic skills are also vital to successful educational outcomes. Many skills can increase students' success in college, but the nature of programming and the role of programming projects in Computer Science make certain types of skills especially critical. We call out, in particular, the skills of time-management, persistence, self-checking, and planning [7]. Studying how students build these skills can help us to guide students to earn academic success.

Our research goal is to discover behavior patterns of CS students in one of those particular skills. Our research mainly focuses on one of the self-checking skills, often called help-seeking behavior. We studied student data in a key CS course at Virginia Tech to find what help-seeking activity leads to academic success.

Help-seeking behavior is an important activity that involves meta-cognitive awareness by students to identify that they have a problem and use self-regulatory skills to receive help [4]. What constitutes help-seeking behavior is slightly controversial, as there is a lot of academic folklore around "good" forms of help-seeking (seeking out help from a GTA to get unstuck from a programming problem) and "bad" forms (depending on constant interaction with the GTA as the only mechanism for making any progress on the programming project). Help-seeking is known to relate to self-regulation, and data on help-seeking behavior can shed light on persistence and self-checking [7]. Self-regulated learning and self-checking are important

behaviors in education and psychology. Significant numbers of studies have found a relationship between academic output and self-regulated learning behaviors among students [20, 28].

We focused on CS students' help-seeking behavior in our department's second semester programming course, CS 2114. Help-seeking behavior in this course typically includes online searching, coming to office hours for instructional staff help, asking questions publicly on online forums, and informal interactions with peers. However, some students do not effectively utilize these resources [25]. We hope that by studying help-seeking behavior, we can discover which patterns lead students to successful outcomes. We hope to discover ways to lead students to utilize better resources in the course for them to maintain academic success.

Our research is based on CS 2114 at Virginia Tech. CS 2114 is the second semester programming course at Virginia Tech. It introduces students to fundamental concepts of object-oriented design and mechanics of data structures. Students in CS 2114 will learn design ideas and strategies to solve programming problems. By analyzing the student data in CS 2114 course, we hope to discover the relationship between help-seeking activities and academic performance.

Our team is a group of five researchers: Dr. Shaffer who is my advisor, Professor Ellis, Ph.D. student Molly Domino and Ph.D. student Patrick Sullivan. Our team chose to collect and analyze CS 2114 course data because it is a mandatory course for CS students. I was a course GTA in the Spring 2021 semester so I was familiar with the course and student behavior. IRB #17-1095 approved our research on studying student behavior in CS 2114. CS 2114 is conducted with one group project (2-3 students), four individual projects, fourteen labs, and two exams. This thesis mainly focuses on the last individual project, Project 4, which is known to be the most challenging individual project overall. All the lab sessions, office

hours, and classes were online during the pandemic. Since all interactions were online, this made it easier to log student interaction data during office hours.

The following research questions guide this work:

- RQ1 Is there a relationship between help-seeking activity and performance?
- RQ2 Is there a difference between how high performers and low performers engage in help-seeking activities?
- RQ3 Do high performers engage more in help-seeking activity than low performers?
- RQ4 How can we identify struggling students?
- RQ5 Is there a difference between struggling and non-struggling students' help-seeking activities?

We decided to divide students in CS 2114 by their Project 4 grade. If a student received higher than 80 on the project, we consider the student to be High Performer and if not, the student would be considered as a Low Performer. We looked at High Performers' and Low Performers' data to find any differences in help-seeking behavior but we could not find any significant differences. We recognize that the High Performers are actually comprised with two groups. Those who did not struggle with the project (and so did not need help) and those who struggled but still did well. When we only considered students who can benefit from help we were able to see the impact of help-seeking behavior on academic performance. The description of how we selected which students struggled on Project 4 is described in Chapter 4.

The rest of the thesis is organized as follow. Chapter 2 covers some background information, primarily describing the meta-cognitive effect on student's academic performance such as self-regulation, help-seeking, self-efficacy, and cognitive load theory. Chapter 3 presents our methodology for collecting necessary data. Chapter 4 describes the analysis of the log data and survey data. Chapter 5 describes and discusses the outcome of the data analysis. Chapter 6 presents ideas for future work. Finally, Chapter 7 presents conclusions.

Chapter 2

Related Work

Meta-cognitive awareness is the awareness of understanding the problem and utilizing the solving process to approach the problem, which can foster students' learning ability. Meta-cognitive awareness is hard to find in novice programmers [21]. But is easily found in high-performing students [8]. Prior research has found that explicit teaching about awareness skills improved students' learning, self-regulation, and growth mindset [15]. Students who experienced explicit teaching of meta-cognitive awareness were also able to explain why their self-efficacy increased and possibly why their performance increased by understanding their process and performance effectively [15]. Another study showed implicit Meta-cognitive skills are essential in students and have a positive effect on CS students' academic performance [2].

Self-regulation is a significant skill in meta-cognition that brings out productive outcomes. Self-regulation is the ability to understand and manage self-behavior. Having disciplined self-regulatory skills can foster learners to productively use newly acquired programming knowledge [14]. Self-regulatory actions can bring benefits to students who want to obtain a successful outcome in the course. However, a previous study has proved self-regulation behaviors are infrequently displayed by novice programmers such as students in introductory CS courses who are often ineffective in reducing programming errors [14]. Loksa et al. showed that self-regulation is effective when a student has adequate prior knowledge, and showed the importance of timing when to teach students to build self-regulated learning. As

self-regulatory skills are crucial in the learning environment, many researchers and instructors are concerned about students under utilizing these self-regulatory skills [25]. To foster students in building stronger self-regulatory skills, it is crucial to study students' behavior. Prior to the Covid-19 pandemic, it was hard to acquire student self-regulatory interaction data because help resources like office hours were held in person and no records were kept. As the courses and office hours were held virtually during and after the pandemic, online systems like Zoom let us collect necessary data to study help-seeking behavior. When most help-seeking related sources are online, it is easier to log students' behavior and conduct research based on the collected log data. Previous research from Harvard tried to investigate the results of virtual office hours on how effective it is. Malan et al. [17] ran virtual office hours in CS 50 course at Harvard and found that virtual office hours were a net positive compared to face-to-face office hours. We hope to utilize our virtual office hours, and online forum log data can allow us to find the relationship between help-seeking behavior and academic performance.

Providing students support and feedback plays a large role in learning a new subject [9]. Even automated feedback can be effective in improving students' academic performance [18]. Automated feedback is the feedback that is given to students by an auto-grader after they submit or run their code. Submission tools like Web-CAT manages automated feedback to students after they submit their assignments [6]. Students need to be engaged in the support and feedback provided by automated feedback. Students have to be active and spontaneous in searching for help and support, which is called Help-seeking behavior. Help-seeking behavior is a cognitive skill and a set of actions of asking for help when necessary [9] and also a very effective skill in self-regulatory action in building learning habits. When novice students encounter difficulty and seek help from instructional staff, the one-

on-one help brought a positive effect on fostering learning [22]. Students can ask for direct one-to-one help from instructors and TAs, which is considered highly effective in fostering learning [22]. Office hours are also a common way instructors provide help to students with face-to-face instructions. There are other types of help-related sources that are provided to students. Piazza, an online forum, enables students to post questions on relevant topics where both instructors and peers can provide support. Searching for online resources with proper inquiry, students can get thousands of answers from experts [24]. Using help related sources in the course, students can receive extra help, direct answers to their problem, and clarification on course materials.

Multiple studies as listed following found factors that have an impact on help-seeking behaviors, which include the tutor's ability [19], accessibility of help [26], and prior experience [22]. While some factors affect help-seeking behavior, studies also have found some unproductive help-seeking behaviors throughout the course. Samiha et al. [18] had done research on finding unproductive help-seeking behavior which are first, immediate help where students seek help in the beginning or even before starting the project, second, high frequency of seeking help, and lastly, requesting for help even when they have ability to solve problem on their own . By studying these relative skills, researchers were able to find patterns and factors that impact help-seeking activities. This research is simlar to our work, but we focused on the effectiveness of help-seeking behavior in CS undergraduate students related to an intermediate-level programming project.

Academic self-efficacy is another factor that has an impact on students' academic performance [12]. Self-efficacy is a student's self-belief in their own capabilities to organize and accomplish given tasks. Self-efficacy can be a motivation for students in earning academic success [12]. As students experience more repeated failures in the learning process, the more

likely that they will develop low self-efficacy [11]. So it is crucial to provide enough help to students to prevent students from creating low self-efficacy. Students with higher academic performance have higher confidence such that they often seek help from instructors [12]. Research found that students with low self-efficacy tend to seek less help than higher self-efficacy students [9]. Students with higher self-efficacy received higher grades than low self-efficacy students. Therefore, many researchers are trying to develop ways to prevent students from forming low self-efficacy.

One factor that affects both self-efficacy and help-seeking behavior is test anxiety. Students who scored high on test anxiety tend to under-perform on exams [3]. Yang et al. also found that students taking online courses with online help have less anxiety. In addition, online courses encourage students to participate in more self-regulatory behaviors, such as receiving support from instructors [27]. Test anxiety is known to be negatively related to self-efficacy and also has a negative relationship with help-seeking behavior [16]. Maier et al. found that in the research of 150 students, more than half of the students never sought help regarding their test anxiety.

Cognitive load theory suggests that when a person is acquiring new information, it would be processed first in the working memory. Working memory has limited capacity. Those who have prior knowledge are using less of this limited memory capacity, which indicates that cognitive load has a negative impact on self-regulated learning for those who are new to the subject [10]. Cognitive load also suggests that students with less knowledge in the subject will need more time to process the information learned from the course material or to complete assignments [1]. Cognitive load can have an impact on student's performance and their self-regulatory actions. As students have prior knowledge of the subject, they have

more time to understand the material than others who do not. Prior knowledge has an impact on students' self-regulatory actions. Research has found that students who already have high expertise in the subject are more likely to engage in self-regulated learning behavior such as seeking help. In contrast, students with low prior knowledge tend to seek less help [5]. The reason for students with insufficient prior knowledge to seek less help is that they are more likely to give up on the assigned tasks [13]. Along with prior knowledge, the proficiency level also affects help-seeking behavior [10]. Proficiency level can be defined as how fluent a student is with a programming language. For novice students, they will have lower proficiency than others who have experience in programming. As a result, novice learners have a hard time finding answers to the problems on their own [27].

Chapter 3

Methodology

Our current efforts focus on data collected from the Spring 2021 and Fall 2021 offerings of a second-semester programming course. We want to analyze the data collected to find the relationship between help-seeking activities and academic performance of CS 2114 students. Our work centered on Project 4 because it is the last project students work on individually and is the most challenging project in CS 2114. As the project is more challenging, the variation in perceiving how difficult the project is will be higher. Due to the COVID-19 pandemic, CS 2114 course was taught in a way that is different from the general historical practice. In particular, the lectures, labs, and office hours were all online, whereas prior to the pandemic, they would have all been given face-to-face. Historically, access to course personnel (instructors or TAs) was in person, via email, or through the course forum (Piazza). The pandemic changed in-person access to online access through Zoom. There was no organized information for CS 2114 collection related to TA or instructor in-person office hours prior to the pandemic. Another source of information for students is informal face-to-face discussions with peers and classmates, often available in our department's large undergraduate lab area. This source of information was also lost during the pandemic. From a data collection standpoint, there were some gains during the pandemic, which are logs of students' actions throughout the semester. As office hours and courses were taught online, we were able to acquire logs of receiving help through office hours and the online course forum Piazza. This section will discuss how we collected necessary data and how we managed and organized the

collected data.

3.1 Log-Based Data Collection

3.1.1 Data collection from Canvas LMS / Web-CAT

Canvas is the Learning Management System (LMS) used by students and instructors at Virginia Tech. Canvas contains students' information on tasks such as assignments, labs, and exams. Using IRB #17-1095 we could gain students' performance data from Canvas LMS. Another method to receive students' help-seeking interaction and performance data was through Web-CAT. Web-CAT is an open-source automated grading system. Web-CAT allows students to upload their programming assignments and receive grades with feedback. The feedback provides general guidelines to students so they can debug or find mistakes within their programs. Web-CAT data not only contains students' performance but also contains other important information for our research such as submission numbers, correctness of the submission, date of the submission, style grade, and design grade. Web-CAT data example is shown in Figure 3.1. We have deleted selected students' information such as name and email for privacy protection.

As we wanted to study students' interaction behaviors in solving Project 4, we had to figure out which category to use from the raw Web-CAT data. We decided to use the following categories:

- Project 4 grade
- Number of submissions
- Date of the submission

Assignment	User	Email	Last Name	First Name	Sub No.	Time	Timestamp	Correctness/	Correctness/	Style Score	Style %	Design	Design %	Penalty/Bon	Total Score	TG
Project 4:	Space Colonies				8	2021-04-09 (	1.6179E+12	30.0289855	0.85797101	15	1	50	1	0	95.0289855	95.0289855
Project 4:	Space Colonies				4	2021-04-15 (	1.6185E+12	22.2964209	0.6370406	15	1	45	0.9	-30	52.2964209	82.2964209
Project 4:	Space Colonies				4	2021-04-13 (	1.6183E+12	27.7641808	0.79326231	15	1	46	0.92	-20	68.7641808	88.7641808
Project 4:	Space Colonies				7	2021-04-12 (	1.6182E+12	26.8195266	0.76627219	15	1	49	0.98	-15	75.8195266	90.8195266
Project 4:	Space Colonies				5	2021-04-06 1	1.6177E+12	31.9390305	0.91254373	13	0.86666667	40	0.8	0	84.9390305	84.9390305
Project 4:	Space Colonies				3	2021-04-09 (	1.6179E+12	0	0	0	0	45	0.9	0	45	45
Project 4:	Space Colonies				3	2021-04-11 (	1.6181E+12	0	0	15	1	44	0.88	-10	49	59
Project 4:	Space Colonies				8	2021-04-08 1	1.6179E+12	16.3394331	0.46684095	15	1	28	0.56	0	59.3394331	59.3394331
Project 4:	Space Colonies				4	2021-04-09 (	1.6179E+12	19.6988432	0.56282409	12	0.8	42	0.84	0	73.6988432	73.6988432
Project 4:	Space Colonies				6	2021-04-08 2	1.6179E+12	26.3179348	0.75194099	15	1	48	0.96	0	89.3179348	89.3179348
Project 4:	Space Colonies				22	2021-04-11 (	1.6181E+12	33.4140911	0.95468832	15	1	50	1	-15	83.4140911	98.4140911
Project 4:	Space Colonies				13	2021-04-16 (	1.6185E+12	25.5837137	0.73096325	11	0.73333333	48	0.96	-35	49.5837137	84.5837137
Project 4:	Space Colonies				4	2021-04-14 (	1.6184E+12	28.9031621	0.82580463	15	1	30	0.6	-25	48.9031621	73.9031621
Project 4:	Space Colonies				7	2021-04-07 (	1.6178E+12	35	1	15	1	50	1	0	100	100
Project 4:	Space Colonies				11	2021-04-12 (	1.6182E+12	22.6086957	0.64596273	14	0.93333333	30	0.6	-15	51.6086957	66.6086957
Project 4:	Space Colonies				5	2021-04-09 (	1.6179E+12	31.9862764	0.91389361	15	1	30	0.6	0	76.9862764	76.9862764
Project 4:	Space Colonies				18	2021-04-08 (	1.6179E+12	20.3490092	0.58140026	15	1	34	0.68	0	69.3490092	69.3490092
Project 4:	Space Colonies				14	2021-04-10 (	1.618E+12	33.8363171	0.96675192	15	1	48	0.96	-5	91.8363171	96.8363171
Project 4:	Space Colonies				11	2021-04-10 (	1.618E+12	27.610786	0.7888796	15	1	30	0.6	-5	67.610786	72.610786
Project 4:	Space Colonies				3	2021-04-10 2	1.6181E+12	0	0	0	0	28	0.56	-10	18	28
Project 4:	Space Colonies				10	2021-04-13 (	1.6183E+12	33.6708861	0.96202532	12	0.8	50	1	-20	75.6708861	95.6708861
Project 4:	Space Colonies				18	2021-04-11 1	1.6182E+12	14.5652174	0.41614907	15	1	30	0.6	-15	44.5652174	59.5652174
Project 4:	Space Colonies				11	2021-04-09 (	1.6179E+12	35	1	15	1	48	0.96	0	98	98
Project 4:	Space Colonies				11	2021-04-09 (	1.6179E+12	28.8848631	0.8252818	15	1	50	1	0	93.8848631	93.8848631
Project 4:	Space Colonies				6	2021-04-16 (	1.6185E+12	24.8136646	0.70896185	15	1	42	0.84	-35	46.8136646	81.8136646
Project 4:	Space Colonies				13	2021-04-10 (	1.618E+12	33.8709677	0.96774194	15	1	42	0.84	-5	85.8709677	90.8709677
Project 4:	Space Colonies				10	2021-04-09 (	1.6179E+12	29.5590478	0.84454422	15	1	48	0.96	0	92.5590478	92.5590478
Project 4:	Space Colonies				5	2021-04-11 (	1.6181E+12	21.4642098	0.61326314	5	0.33333333	44	0.88	-10	60.4642098	70.4642098
Project 4:	Space Colonies				20	2021-04-01 1	1.6173E+12	35	1	15	1	50	1	0	100	100
Project 4:	Space Colonies				9	2021-04-09 (	1.6179E+12	33.2438069	0.94982305	15	1	50	1	0	98.2438069	98.2438069
Project 4:	Space Colonies				4	2021-04-14 (	1.6184E+12	0	0	0	0	35	0.7	-25	10	35
Project 4:	Space Colonies				10	2021-04-10 (	1.618E+12	7.98418972	0.22811971	15	1	30	0.6	-5	47.9841897	52.9841897
Project 4:	Space Colonies				31	2021-04-10 (	1.618E+12	13.4291329	0.38368951	15	1	35	0.7	-5	58.4291329	63.4291329
Project 4:	Space Colonies				14	2021-04-09 2	1.618E+12	24.7163723	0.70618207	15	1	35	0.7	-5	69.7163723	74.7163723
Project 4:	Space Colonies				11	2021-04-02 1	1.6174E+12	35	1	15	1	50	1	0	100	100
Project 4:	Space Colonies				9	2021-04-11 (	1.6181E+12	32.1532984	0.91866567	15	1	48	0.96	-10	85.1532984	95.1532984
Project 4:	Space Colonies				4	2021-04-10 (	1.618E+12	23.6231884	0.67494824	13	0.86666667	35	0.7	-5	66.6231884	71.6231884
Project 4:	Space Colonies				14	2021-04-09 1	1.618E+12	22.7187848	0.64910814	15	1	44	0.88	-5	76.7187848	81.7187848

Figure 3.1: Example of Web-CAT student data logs

- Correctness of the submission
- Penalty (if any applies)

Project 4 grade data are used to examine the academic performance, and rest of the categories are used to determination factors to divide students into groups.

3.1.2 Data collected from Office hours

As the course and office hours were managed virtually, we could gain the logs of students' visits to the course office hours. Along with the logs from Zoom sessions, instructors and GTAs of CS 2114 managed a sign up log sheet for students. Students were required to sign up to get help from GTAs during online office hours. Students had to sign in to the sign up sheet with the reason for their visits and specific TA they wanted to receive help from. Zoom logs contained time spent in the TA office hours zoom sessions (during the pandemic,

students had to sign up on a Google doc in advance for access to a TA via Zoom, whereas this was not tracked at all previous to the pandemic). Canvas LMS contained logs of other activity on the course forum, and project and lab grades.

With online office hours, we were able to obtain Zoom visit logs. We collected both Zoom visit logs and sign up sheets after the project due date, including the grace period for late submissions. While the sign up sheet offered some valuable insights on the patterns of student help-seeking behavior, it also contained some missing or extraneous data. We were able to verify the missing or extra data by comparing the sign up sheet and the Zoom visit logs. We confirmed that some students did not sign in but still came to the office hour, and some just decided not to sign in but joined anyways during the office hour for help. TAs from CS2114, including myself, confirmed that there were a few students who deleted others' names in the sign up sheet so they could visit the TA faster. Later on, we changed the type of the sign up sheet from Google document to Google Forms so that students could not modify the sign up sheet.

Figure 3.2 is an example of the initial sign up sheet TAs maintained. In the fifth column, each student had to write the reason for coming to the office hour. We first had to sort out which students came in during the office hour with Project 4 questions. We sorted out students who signed up for Project 4 questions which came to 927 student sign up data entries. While comparing the sign up sheet with Zoom, we found students who came back on the same day without signing up or were not signed up at all. Therefore, we concluded that the sign up sheet had some missing data, and we used the Zoom log as the basis for office hour data collection.

We used Zoom visit logs for the base data and compared the log with the filtered sign up sheet. The time frame is from 3/27/2021 which is the Project 4 release date to 4/15/2021

Timestamp	TA will mark	Please enter your Name (zoom)	Enter a brief description of your question	Specific TA you want to receive help	Additional Co
4/15/21 9:17	x		Question about Project 5		
4/15/21 10:57	x		My project 5 group		
4/15/21 11:09	x		P5 Design + Intermediate Sub		No
4/15/21 11:12	x		Project 5		
4/15/21 11:18	x		project 4 colony calculator		
4/15/21 11:19	x		Project 4 colony cal		no
4/15/21 11:19	x		Project 5 UML diagram		
4/15/21 11:20	x		project 4 colony cal		no
4/15/21 11:21	x		Testing stuff		Yerrrrr!!!
4/15/21 11:37	x		P5 Intermediate Submis		
4/15/21 12:11	x		project 4		
4/15/21 12:58	x		I have a question regarding the group from project 5		
4/15/21 12:59	x		Lab 13, size and toPostOrderString implementation		
4/15/21 13:56	x		Lab 13, height method StackOverFlow error		
4/15/21 16:59	x		Proj 4 accept		will be back sc
4/15/21 14:32	x		Project 4		
4/15/21 15:00	x		project 4 colonyreader		No
4/15/21 15:26	x		Colony Calculator- P4		thank youu
4/15/21 15:28	x		Project 4		
4/15/21 15:31			Required Lab Hours for Project 5		
4/15/21 16:15	x		Project 4- Calculator		
4/15/21 16:44	x		Project 4 Compilation Errors		
4/15/21 16:47	x		Project 4		
4/15/21 16:55	x		project 4 compare to		no
4/15/21 17:10	x		Project 5 Group Meeting		
4/15/21 17:25	x		null pointer in ColonyReader		
4/15/21 17:48	x		Project 4		Im sorry Im ba
4/15/21 17:51	x		project 4 SpaceColonyDataException		
4/15/21 17:51	x		Applicant Issue, planet for person?		Thank you, I'n
4/15/21 18:11	x		Project 4		
4/15/21 18:29	x		come on in to the break out room ill give you the link to your lab		

Figure 3.2: Example of Student sign up sheet for Office hour

which is the last day to submit the project with penalty. Zoom visit logs track every student's activity, such as entering the waiting room, moving to the main room, and moving to a breakout room. Unfortunately, the Zoom log does not explicitly label which room a student is in, just that some transition occurred. Therefore, we needed to pre-process the Zoom log. The raw Zoom data contained unnecessary data like students coming for other assignments than Project 4 or students who went into the waiting room but were unable to see TA for whatever reasons. We decided to take three steps to sort out who actually came to the office hours for Project 4 and saw a TA for help. First, we deleted students who came in asking for help with other assignments than Project 4 by comparing the Zoom log with initial sign up sheet. Then, we had to find students who actually received support from

office hours on Project 4. We confirmed that Zoom logs contain students moving from one room to another room, and we have established that when a student comes into Zoom, they log students entering the waiting room, main room, and then a breakout room where a TA is. However, there was another problem with the Zoom log. For some students, there were three logs showing that the student came into office hours (waiting room, main room, and then breakout room), whereas, for some students, there were only one or two log entries. By comparing the Zoom log with the sign up sheet, we were able to find that the visit logs of less than 3 minutes mean that the student was moving around the room or even left without seeing the TA. Another factor we needed to sort out was students who could not see TAs. We went through the data to see if there were any students who only had one log in Zoom, which meant that they only came into the waiting room. After going through these steps, we were able to find the exact logs of students who actually visited the TA to receive help in Project 4. There were 2979 log entries in original data of two weeks of Zoom session, and after pre-processing, we had 1677 visit logs. There was 177 visit logs of students from 4/7/2021 to 4/16/2021 of students who entered the Zoom waiting room but unable to progress into the main room to receive help. Therefore, the final count of the Zoom visit log was 1500 visit logs. The items from the zoom visit logs that we collected are following:

- Name
- User Email
- Join Time
- Leave Time
- Duration

Figure 3.3 shows an example of the Zoom log that we collected. We deleted student's personal information such as name and email from the data for privacy protection.

Name (Original Name)	User Email	Join Time	Leave Time	Duration (Minutes)
		4/14/21 17:24	4/14/21 17:32	8
		4/8/21 18:09	4/8/21 19:17	69
		4/8/21 19:20	4/8/21 20:23	63
		4/8/21 11:01	4/8/21 14:45	225
		4/8/21 14:45	4/8/21 15:52	68
		4/14/21 19:02	4/14/21 19:05	4
		3/29/21 13:32	3/29/21 13:52	20
		3/29/21 13:52	3/29/21 14:00	8
		3/29/21 14:33	3/29/21 14:45	12
		3/29/21 14:45	3/29/21 14:52	8
		3/30/21 15:32	3/30/21 15:39	7
		3/30/21 15:39	3/30/21 15:50	11
		3/30/21 18:48	3/30/21 18:53	5
		4/1/21 20:39	4/1/21 20:46	7
		4/2/21 10:10	4/2/21 10:18	9
		4/3/21 13:42	4/3/21 13:51	9
		4/5/21 11:32	4/5/21 11:47	16
		4/5/21 11:48	4/5/21 11:55	8
		4/5/21 12:22	4/5/21 12:31	9
		4/5/21 13:05	4/5/21 13:09	5
		4/5/21 13:09	4/5/21 13:20	11
		4/5/21 14:10	4/5/21 14:53	44
		4/5/21 14:53	4/5/21 14:56	4
		4/5/21 17:52	4/5/21 18:15	23
		4/5/21 18:15	4/5/21 18:43	29
		4/7/21 11:59	4/7/21 14:22	143
		4/7/21 14:22	4/7/21 14:30	8
		4/7/21 14:48	4/7/21 18:10	203
		4/7/21 18:10	4/7/21 18:19	9
		4/7/21 18:38	4/7/21 19:16	39
		4/8/21 16:05	4/8/21 16:23	19
		4/8/21 16:23	4/8/21 16:31	8
		4/8/21 18:17	4/8/21 19:08	52
		4/12/21 11:10	4/12/21 11:49	39
		4/12/21 11:49	4/12/21 12:01	12
		4/12/21 13:08	4/12/21 13:52	44
		4/12/21 13:52	4/12/21 14:13	21
		4/13/21 12:05	4/13/21 12:17	12

Figure 3.3: Example of Student Office hour Interaction Zoom data

By having the zoom log, we could calculate the number of each students' visits to the office

hour. We hoped that these categories would enable us to find the patterns of help-seeking behavior of each student, which will bring out the outcome of how these help-seeking behavior patterns lead students to success. First, we sent out consent forms to students and asked if they agreed to let us collect demographic data and survey data. Then we collected grades and penalties students received through Canvas, collected zoom logs of office hour visits, and lastly, the sign up log of office hour visits.

3.1.3 Data Cleansing

Our raw student grade data collection had 569 students and we have found that 51 students did not submit the project. For students who received lower than 25 on the project, we considered them to have given up on Project 4 or not seriously attempted the project. We found 25 students who received less than 25 on the project and we decided to filter them out from our data set. We then compared the Web-CAT data and Canvas LMS data to see if the records matched. After filtering out unnecessary data, we had data for 488 students. Figure 3.4 shows the data collected for log-based data collection.

3.2 Data Collection from Piazza

Piazza is an online forum where students can post questions to each other and instructional staff. Instructors can set folders of assignments on Piazza so students can categorize their questions as related to those assignments. Piazza provides several statistics including Activity Breakdown Report, Post Views, Participation by Folders that instructors create, Participation by Threads, and lastly, Poll Statistics. Within the following statistics categories, we used Participation by Folders. As our research is based on Project 4, we collected

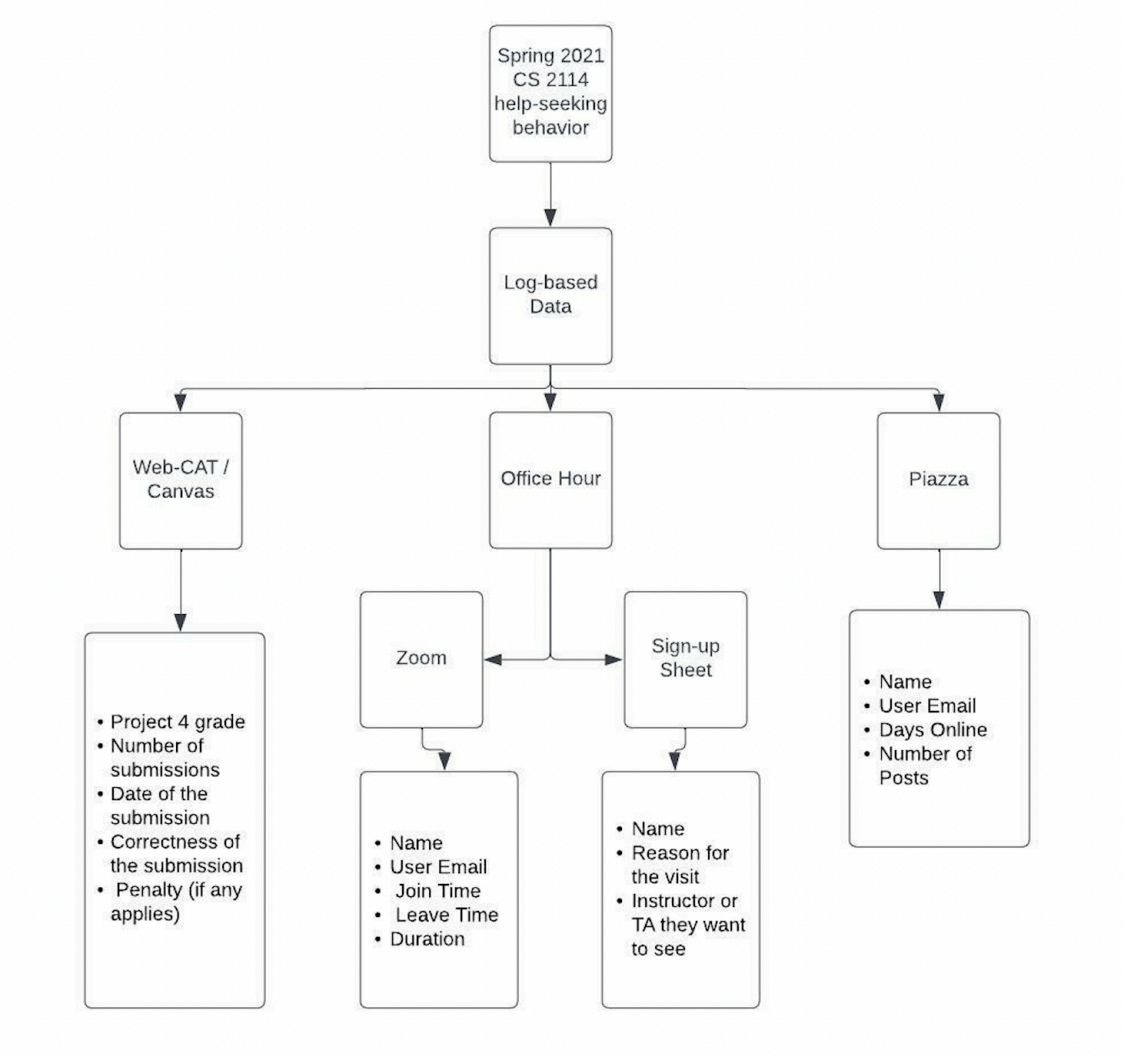

Figure 3.4: Visualization on necessary Log-based Data Collection

participation statistics on the Project 4 folder. Extracted data from Piazza contains student names, emails, roles, the number of days students were online within the timeline of Project 4, the number of posts students posted, the number of answers students posted, edits to the answer they wrote, follow-ups, and lastly replies to follow-ups. Figure 3.5 shows an example of the Piazza data. We deleted student names and emails for privacy protection. Within the extracted data, we used the following categories to find relationship between students' Piazza activities and academic performance:

- Name
- User Email
- Days Online during Project 4 time frame
- Number of Posts

name	emails	role	days online	posts	answers	edits to answ	followups	replies to foll
		student	25	0	0	0	0	0
		student	90	0	0	0	0	0
		student	21	0	0	0	0	0
		student	12	0	0	0	0	0
		student	39	0	0	0	0	0
		student	19	0	0	0	0	0
		student	26	0	0	0	0	0
		student	117	0	0	0	0	0
		student	87	0	0	0	0	0
		student	71	0	0	0	0	0
		student	33	0	0	0	0	0
		student	2	0	0	0	0	0
		student	38	0	0	0	0	0
		student	114	0	0	0	1	0
		student	12	0	0	0	0	0
		student	68	0	0	0	0	0
		student	132	0	0	0	0	0
		student	6	0	0	0	0	0
		student	89	5	0	0	5	1
		student	44	0	0	0	0	0
		student	23	1	0	0	1	1
		student	33	0	0	0	0	0
		student	86	0	0	0	0	0
		student	39	0	0	0	0	0
		student	85	0	0	0	0	0
		student	165	1	0	0	0	0
		student	85	5	0	0	4	13
		student	130	0	0	0	0	0
		student	12	0	0	0	0	0
		student	75	0	0	0	0	0
		student	195	3	0	0	0	0
		student	10	0	0	0	0	0
		student	71	0	0	0	0	0
		student	24	0	0	0	0	0
		student	81	0	0	0	0	0
		student	58	1	0	0	0	0
		student	47	0	0	0	0	0

Figure 3.5: Example of Piazza data on Project 4

3.2.1 Data Collection

There were 141 threads marked as Project 4 related and 36 threads marked as project but related to Project 4 where 7 threads were related to Project 4 Web-CAT submission. We exported raw piazza data of 569 students by using 170 threads that are related to Project 4. By clustering the Piazza data with Project 4 grade data, we were able to confirm that 51 students did not submit the project and 12 students did not use Piazza. As 12 students did not use Piazza at any time of the Project 4, we counted 0 for all categories of number of days online, number of posts, and number of answers. We also subtracted students who received less than 25 on the project, considering them as students who did not seriously attempt the project and gave up on the project. At the end of data collection, we had 488 student Piazza interaction data which was the same as the office hour interaction data. The 488 student data entries had names, emails, number of days online during Project 4, number of questions related to Project 4 posted on Piazza, number of post views related to Project 4 on Piazza and number of answers they posted related to Project 4 on Piazza.

3.3 Finalized Help-seeking Data

After collecting office hour logs and Piazza logs, we linked them together. The linked data had 488 students with data on project performance, correctness of the project by third submission, number of days spent on project, maximum number of submissions, number of visits to the office hours, number of questions posted on Piazza, number of post views on Piazza, and number of answers each student answered on Piazza. Figure 3.6 is an example of the final help-seeking data.

Course	User	Last Name	First Name	Email	View	days online	posts	answers	NumVisit	StartDate	NumberOfDa	Max Sub	Time	Correctness I	Penalty/Bon	Grade	Total Score
CS 2114					0	0	0	0	0	4/9/21	1	1	4/9/21	0	0	28	28
CS 2114					0	12	0	0	0	4/13/21	1	1	4/13/21	0	-25	45.1136045	20.1136045
CS 2114					0	19	0	0	0	4/12/21	1	1	4/12/21	0	-15	49	34
CS 2114					0	6	0	0	0	4/14/21	1	1	4/14/21	0	-30	96.1956522	66.1956522
CS 2114					13	62	0	0	0	4/16/21	1	1	4/16/21	0	-35	47.2486332	12.2486332
CS 2114					0	47	0	0	0	4/16/21	1	2	4/16/21	0	-35	35	0
CS 2114					0	14	0	0	1	4/8/21	2	2	4/8/21	0	-15	28	0
CS 2114					1	19	0	0	0	4/9/21	1	2	4/9/21	0	-5	29.5606104	0
CS 2114					2	9	0	0	0	4/9/21	1	2	4/9/21	0	-5	25	0
CS 2114					4	47	0	0	0	4/13/21	1	2	4/13/21	0	-25	85.4468459	6.96908771
CS 2114					11	71	0	0	4	4/11/21	1	2	4/11/21	0.81270903	-10	42	0
CS 2114					34	66	0	0	0	4/9/21	1	2	4/9/21	0	0	76.6339079	28.6339079
CS 2114					67	27	0	0	0	4/9/21	1	2	4/9/21	0.77217391	-5	30	0
CS 2114					0	14	0	0	0	4/8/21	1	3	4/8/21	0.16284182	0	35.6994636	0
CS 2114					0	17	0	0	0	4/16/21	1	3	4/16/21	0.10180276	-35	38.5630965	0
CS 2114					0	17	0	0	0	4/10/21	1	3	4/10/21	0.2613354	-5	44.1467391	4.04651535
CS 2114					0	21	0	0	0	4/15/21	1	3	4/15/21	0.77466343	-35	77.11322	0
CS 2114					0	21	0	0	1	4/10/21	2	3	4/10/21	0	-10	59	0
CS 2114					0	8	0	0	2	4/11/21	1	3	4/11/21	0.94586813	-10	98.1053844	23.1053844
CS 2114					1	25	0	0	0	4/16/21	1	3	4/16/21	0	-35	40	0
CS 2114					1	13	0	0	0	4/9/21	2	3	4/9/21	0	-20	54	0
CS 2114					1	23	0	0	0	4/16/21	1	3	4/16/21	0.95767196	-35	83.5185185	0
CS 2114					1	33	0	0	1	4/8/21	2	3	4/8/21	0.66788895	0	70.3761132	21.1351211
CS 2114					2	36	0	0	0	4/16/21	1	3	4/16/21	0	-35	45	0
CS 2114					2	11	0	0	0	4/9/21	1	3	4/9/21	0	0	45	0
CS 2114					2	40	0	0	0	4/7/21	1	3	4/7/21	0.82497213	0	84.8740245	0
CS 2114					2	41	0	0	7	4/9/21	2	3	4/9/21	0	-10	28	0
CS 2114					3	44	0	0	0	4/12/21	2	3	4/12/21	0.39340814	-25	48.7692847	0
CS 2114					3	83	0	0	8	4/16/21	1	3	4/16/21	0	-35	50	0
CS 2114					4	9	0	0	4	4/10/21	1	3	4/10/21	0	-5	43	0
CS 2114					4	37	0	0	16	4/9/21	1	3	4/9/21	0.74192373	-5	25.9673306	20.9673306
CS 2114					6	19	0	0	0	4/7/21	1	3	4/7/21	1	0	100	0
CS 2114					8	39	0	0	0	4/12/21	1	3	4/12/21	0.94543332	-15	98.0901663	16.9054567
CS 2114					8	73	1	0	13	4/10/21	1	3	4/10/21	0.8844051	-5	95.9541786	22.8195782
CS 2114					9	46	0	0	0	4/9/21	2	3	4/9/21	0.81069403	-5	78.3742911	23.2759092
CS 2114					10	68	2	0	0	4/9/21	1	3	4/9/21	0.85007496	0	91.7526237	29.9127907
CS 2114					35	50	0	0	0	4/9/21	1	3	4/9/21	0.82689211	0	91.9412238	26.0046308

Figure 3.6: Final help-seeking daa

3.4 Survey Data collection

Survey data was collected from the second semester of the CS2 course (Fall 2021). Our team had several discussions on which questions to collect. The main reason for the survey is to find out the percentage of those who found the project challenging and how they managed the difficulties, for example, receiving help. Therefore, the inventory included questions like the difficulty of the project, the grade the student received, how much support the students received if they had attended all the lectures, and more in order to understand students' responses toward project 4. The survey questions were designed as multiple-choice questions except the last question which was to write any feedback on Project 4.

Survey questions are based on Project 4 and how students managed themselves with the project. These questions allow us to confirm the percentage of students who thought the project was difficult in the previous semester and how students used their self-regulation to solve Project 4. 347 of 456 students (76%) agreed to participate in the study. The survey

was taken anonymously, therefore it included questions on the score for Project 4 and how difficult each student perceived the project to be.

Questions are listed below:

Q4 How often did you attend class lectures?

①*Only a Few Lectures* ② *Less than Half Lectures* ③ More than Half Lectures ④ *Most of the Lectures* ⑤ *All the Lectures*

Q5 What was your project 4 score?(range)

① *Below 60* ② 70 - 79 ④ 80 - 89 ⑤ Above 90

Q6 Which answer best describes how difficult you found project 4 to be?

① Very difficult ② Difficult ③ Not Difficult ④ Easy

Q7 How often did you use outside source for project 4 (StackOverflow or CSAwesome)

① None ② 5 or less ③ 6 to 10 times ④ ore than 10 times

Q8 How often did you seek help from instructional staff?

① None ② 5 or less ③ 6 to 10 times ④ ore than 10 times

Q9 How often did you seek help from individuals aside from instructional staff?

① None ② 5 or less ③ 6 to 10 times ④ ore than 10 times

Q10 Do you think you had sufficient access to help during project 4 from instructional staff?

① Strongly Disagree ② Disagree ③ Agree ④ Strongly Agree

Q11 Do you think help from instructional staff and online forums like Piazza helped?

① Strongly Disagree ② Disagree ③ Agree ④ Strongly Agree

Q12 Any feedback?

Chapter 4

Spring 2021 Help-Seeking Data Analysis

After collecting help-seeking behavior data (office hours and Piazza), we analyzed the relationship between aspects of the log data and Project grades. Our broader goal is to find the relationship between help-seeking activities and academic performance. Our data analysis seeks to find the connection between student help-seeking behavior and student performance on Project 4. With the collected data, we calculated Spearman's correlation coefficient to find the connection between the behavior and the outcome. The number of visits to the office hour and the project performance we used was not continuous. Our data can be categorized as ordinal, for example, counting the number of visits to office hours. Since Spearman's correlation evaluates two continuous or ordinal variables, we used Spearman's correlation instead of Pearson's to create a correlation based on ranked values. We calculated Spearman's correlation between the number of visits to the office hour and Project grade and the number of questions asked on Piazza and Project score. The outcome is presented in Table 4.1.

	Correlation	R-value	Sig. (P-value)
Number of office hour visit	0.079	0.117	0.00982
Number of question views on Piazza	0.285	0.217	$<$ **0.01**

Table 4.1: Correlation between number of office hour visit and Project scores and number of question views on Piazza and Project scores on original data

The correlation value came out as 0.079, showing almost no correlation between the number of visits to the office hours and the project score. Taken at face value, this result appears to indicate no relationship between help-seeking behaviors and academic performance. Our next attempt to analyze the relationship between help-seeking behavior and project performance is to focus on students who were low performers on Project 4. We start by looking at the score distribution, shown in Figure 4.1

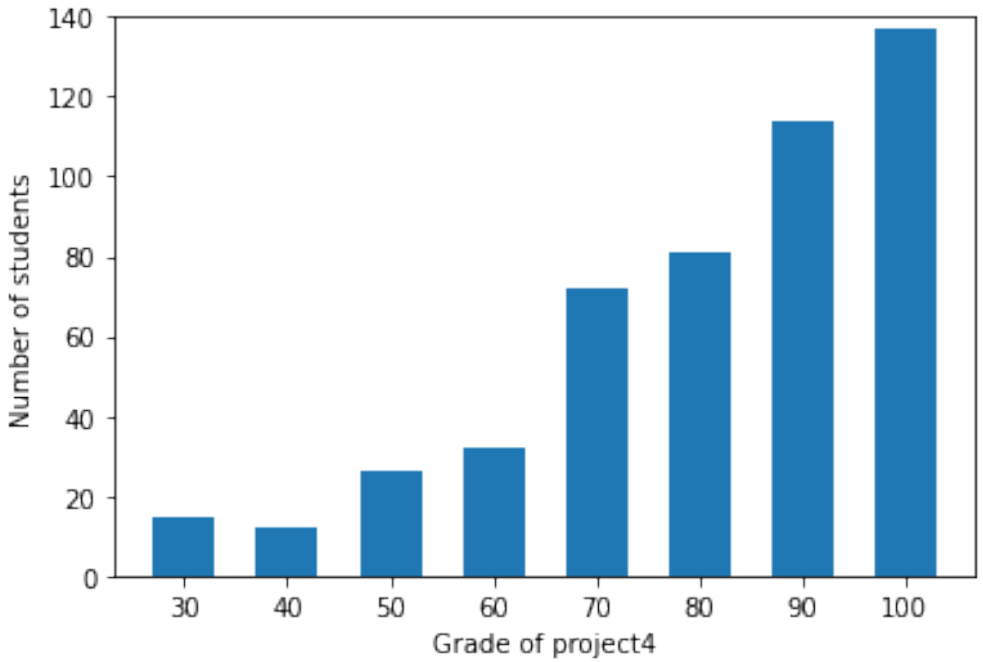

Figure 4.1: Distribution of the project 4 grade

We start this process by deciding who is a high performer on the project versus who is a low performer. We attempt to do this using clustering on the project scores. KMean clustering was used to group students. We used KMean clustering on the scatter plot of the number of visits to office hours versus the project's final grade. Graph 4.2 shows the clustering of the student groups with K = 2. Based on these results, we concluded to use a score of 80 to divide students into two groups.

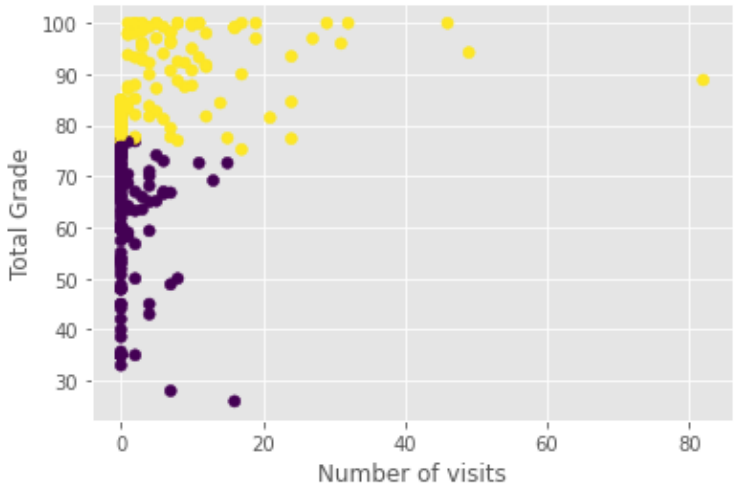

Figure 4.2: Kmean clustering graph

Students who score below 80 on the project are clearly considered to have struggled with the project. We have created Spearman's correlation on LP students to find the relationship between help-seeking behaviors and academic outcomes where the correlation results are represented in Table 4.2

	Correlation	R-value	Sig. (P-value)
Number of office hour visits	-0.005	0.135	0.136
Number of question views on Piazza	0.145	0.062	0.497

Table 4.2: Correlation between number of office hour visits and Project score and number of question views on Piazza and Project score on LP student data

4.1 Help-seeking behaviors in HP and LP students

We believe that analyzing the help-seeking activities of LP will allow us to find the relationship between help-seeking behavior and academic performance. In this section, HP students are who received above 80 on the project, and LP students are who received less than 80

on the project. We wanted to find if there is any difference between HP and LP students in help-seeking behaviors. The first analysis was on the Office hour patterns of HP and LP students, which is shown in Table 4.3. Some factors differ between LP and HP students: average number of visits and duration of the office hour visits. For LP students, the office hour visits' mean duration was 46.01 minutes (with SD = 61.21 minutes). Meanwhile, for HP students, the mean duration of the office hour visits was 36.97 minutes (with SD = 54.80 minutes). As we ran the T-Test with the difference of the mean of the duration of the office hour, the T-Test = -1.6643 and P-value = 0.097, which is higher than our set P-value of 0.05. We could not conclude that HP students' office hour visit duration is longer than LP students'. Another factor was the number of visits to the office hours. For LP students, the mean number of visits to the office hour was 11.66 times (with SD = 7.3) for each student, whereas for HP students, the mean number of visits to the office hour was 34.38 (with SD = 21.05 times) for each student. With T-test = 17.0332 and P-value = $2.209E^{-59}$; therefore, the HP students' number of visits to the office hours is about 3 times more than LP students' and the difference is significant. We could conclude that HP students seek more direct help during office hours than LP students do.

	Project Performance		Duration of office hour visits (Min)		# of visits	
	Mean	SD	Mean	SD	Mean	SD
LP	64.94	17.3486	46.11	61.21	3.368	7.30
HP	95.46	14.5255	39.98	54.80	2.747	4.91
T-test	30.1441		-1.6643		17.03324	
Sig. (P-value)	**< 0.01**		0.097		**< 0.01**	

Table 4.3: Average calculated values to office hour visits and Paired Sample Test on LP and HP students

We ran the correlation between the students' grades, the duration of office hour visits, and the number of times they visited office hours on both HP and LP students. However, the

result came out that both LP and HP student's office hour visits had no relationship with their project grade. The significance of these correlations even came out as 0.1362 and 0.4295 for LP and HP students, shown in the Table 4.4. We were unable to see any relationship between the number of office hour visits and project performance in LP and HP students.

	Average # of visits to office hour	Sig. (P-value)
LP	-0.005	0.13620
HP	0.021	0.4295

Table 4.4: Correlation between Students' Project 4 grade and average number of visits to the office hour by project performance and significance test

4.5.

	# of Post Views		# of Question posted	
	Mean	SD	Mean	SD
LP	0.35	0.58	0.0	0.39
HP	25.311	26.50	0.38	0.1
T-test	-17.9817		-7.295	
Sig. (P-value)	**< 0.01**		**< 0.01**	

Table 4.5: Average number of Piazza activities and Paired Sample test on LP and HP

On the number of the Project 4 posts views, LP students had mean views of 0.35 times (with SD = 0.58), and HP students had the mean views of 25.311 times (with SD = 26.50). The T-test = -17.982 and P-value = $3.18E^{-52}$. For the number of posts on Piazza, LP students' mean number of posts was 0 posts (with SD = 0.39), whereas, for HP students, the mean number of posts was 0.38 posts (with SD = 0.1). The T-test = -7.295 and P-value = $1.88E^{-12}$. As all three categories' calculated P-values came out lower than 0.05, we could conclude that HP students' help-seeking behavior on Piazza and LP students' help-seeking behavior on Piazza had a significant difference. HP students had higher activity patterns of seeking help on Piazza than LP students.

We created correlation between Piazza activities and Project 4 performance to see if there is any relationship. Table 4.6 shows the results of the correlation. For LP students, the correlation between the number of post views on Piazza and the project came out as 0.145 with a significance of 0.4964. For HP students, the correlation came out as 0.160 and a significance of 0.00479. HP student's correlation value was significant, whereas for LP students it was not. However, with a low correlation value, we conclude that there is no relationship between Piazza post views on Project 4 and Project 4 performance.

	Average # of Post Views on Piazza	Sig. (P-value)
LP	0.145	0.4964
HP	0.160	**0.00479**

Table 4.6: Correlation between Students' grade and average number of post views on Piazza for Project 4 by project performance and significance test

As a result, were we could not find any relationship between help-seeking behavior in office hour and piazza with the project outcome.

4.2 Office Hour Data Analysis on Students who Struggled with Project 4

Our previous analysis showed a difference in the patterns of help-seeking activities between HP and LP students, but we could not find any relationship between help-seeking behavior and academic performance in LP students.

After further consideration, we realized that the high-scoring group represents at least two types of behaviors. Some of these students find the project to be relatively easy and so would have no reason to seek help. But since they did not seek help, they appear indistinguishable from struggling students who simply did not seek the help that they needed. Other high-

scoring students did struggle with the project, sought help, and benefited from the help received. Our first question was how can we distinguish students who can benefit from help-seeking behaviors? The answer was students who struggled with the project would be the students who could benefit from utilizing help-seeking behaviors. At the same time, students who would not seriously attempt the project or who perceived the project to be relatively easy would not benefit from using help-seeking resources like office hours and Piazza. After analyzing, we came up with specific categories to identify who might have struggled with Project 4.

4.2.1 Office hour Data Clustering

In Section 5 we discuss the survey data results where 99.1% of LP students said the project was either difficult or very difficult, there were 72.8% of HP students who said the project was difficult or very difficult. As we needed to figure out who struggled among HP students, we started to look at their help-seeking behavior and their Project 4 submission activities, such as maximum submission numbers, the number of days they spent on the project, or the project submission coverage. These were included in the data collection mentioned in Section 3. We came up with ways to sort out who might have struggled and could benefit from help-seeking activities. After through discussion, we decided to use following rules to find who struggled with the Project 4 which are the correctness of the project by third submission and number of submission for Project 4. The average submission activities shown in Table 4.8. Figure 4.3 details our rules for defining struggling students. We were able to find 213 students using this method. We then calculated the correlation of the help-seeking behavior and academic performance. Table 4.7 shows the results.

	Currently used Method
Office Hour correlation	0.313
Piazza Project 4 related post view	0.412

Table 4.7: Difference on calculate correlation on using two different method

After sorting out who struggled with Project 4 within HP and LP, we clustered the two data sets into one data set so we could run an analysis for the outcome. We confirmed that 203 students in the Spring 2021 CS2114 had struggled with the project.

	#
Average of maximum number of submission	10
Average correctness of Project 4 by third submission	48.2%
Average number of days spent on Project 4	2.4 days

Table 4.8: Average of Project 4 submission activities on whole student data

4.2.2 Office Hour Data Analysis Results on Students who Struggled on Project 4

We created Spearman's correlation again for the 213 students that we believe struggled with Project 4. We used Spearman's correlation because extracted office hour data are a non-ordinal category. In this section, we will be discussing the difference in the correlation results on original data and filtered data of students who we believe struggled with Project 4. Figure 4.4 shows the scatter plot of the original data, and Figure 4.5 shows the scatter plot with a linear regression line on students who struggled with Project 4. Figure 4.6 represents the scatter plot with linear regression on students who did not struggle on Project 4. Calculated Linear regression on the office hour data and it's significance in Table 4.9.

- Original Office Hour data linear regression: $y = 0.225x + 85.410$

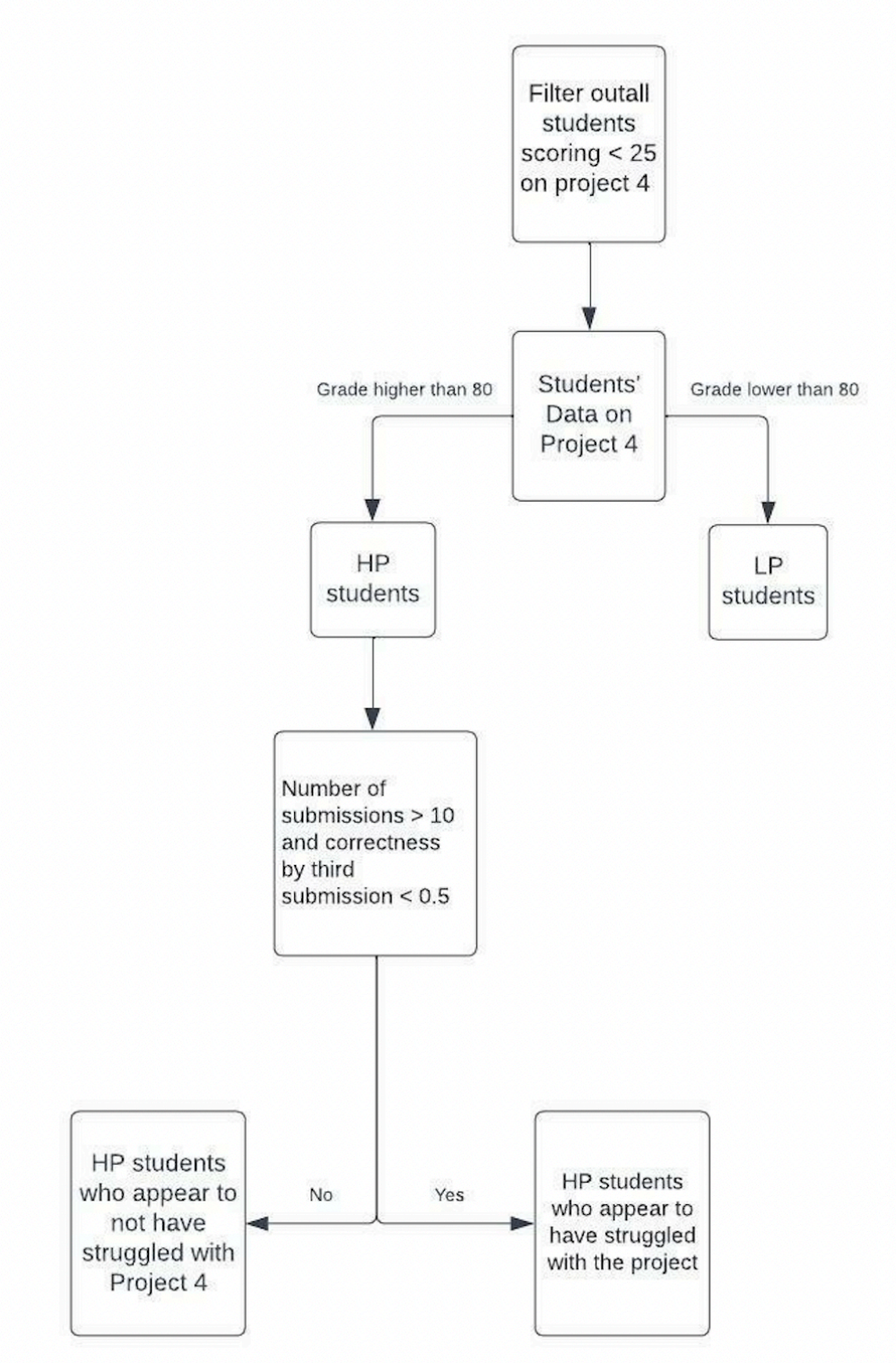

Figure 4.3: Second Data pre-processing algorithm

- Struggling student Office Hour data linear regression: $y = 0.48509x + 75.6514$

- Non-struggling student Office Hour data linear regression: $y = 0.04106x + 93.9492$

	Original Data	Struggling student Data	Non-struggling student Data
Slope	0.225	0.4850956	0.04106
R-value	0.11666	0.270042	0.039304
STD Err	0.08675	0.1137	0.06423
P-value	**0.009817**	**< 0.01**	0.51552

Table 4.9: Office hour Data Linear Regression Test

We then ran linear regression analysis to compare if the difference between scatter plot is significant. Table 4.9 shows the result of the calculated linear regression analysis on the data. The calculated R-value for original data was 0.1167 whereas for struggling student data it was 0.2700 and for non-struggling student it was 0.0393. P-value between original data, struggling student data, and non-struggling student data came out as 0.0098177, $6.53879E^{-05}$, and 0.5155 in order. We can conclude that the difference between the relationship of number of office hour visits and academic performance is moderately high between original data and struggling student data. We can conclude that this difference is significant. We also found that there is no meaningful correlation between performance and help-seeking for non-struggling students, which is confirmation of what we can reasonably expect.

As shown in the scatter plot, Spearman's correlation on struggling student data came out moderately positive, which is 0.543. The Spearmans' correlation on students who we argue to not have struggled with Project 4 came out as 0.007. By observation, the linear regression is much higher than the original one. With a relatively positive correlation value, we can argue that students who struggled with the project can benefit from seeking more help from instructional staff.

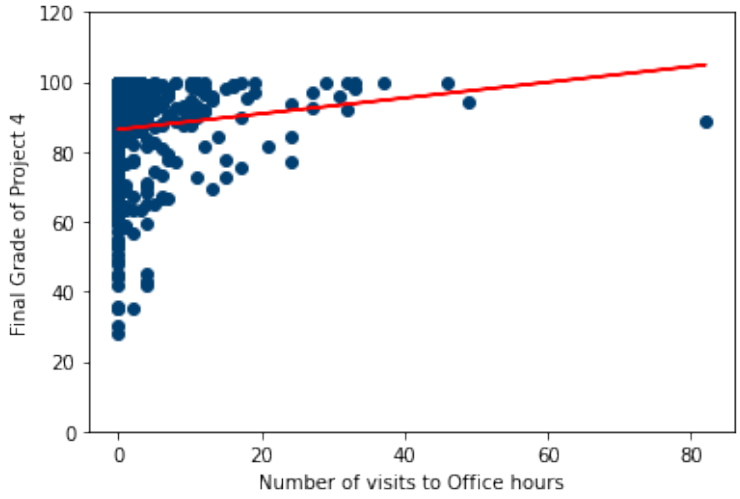

Figure 4.4: Original data scatter plot between number of visits to office hours and final grade of project 4

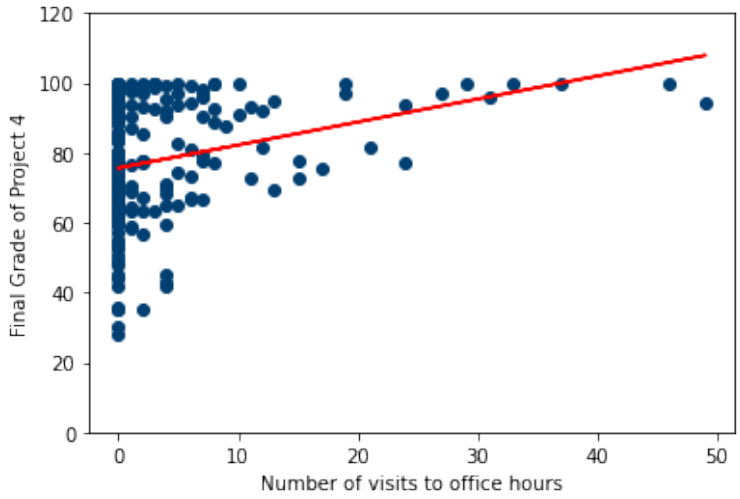

Figure 4.5: Struggling student data scatter plot between number of visits to office hours and final grade of project 4

4.2.3 Zoom Log Analysis between Students who struggled and who did not on Project 4

In the Spring 2021 CS2 course, 213 students came into office hours for direct help from instructors and TAs and the total office hour visit log is 1677 logs. A total of 41.6% of

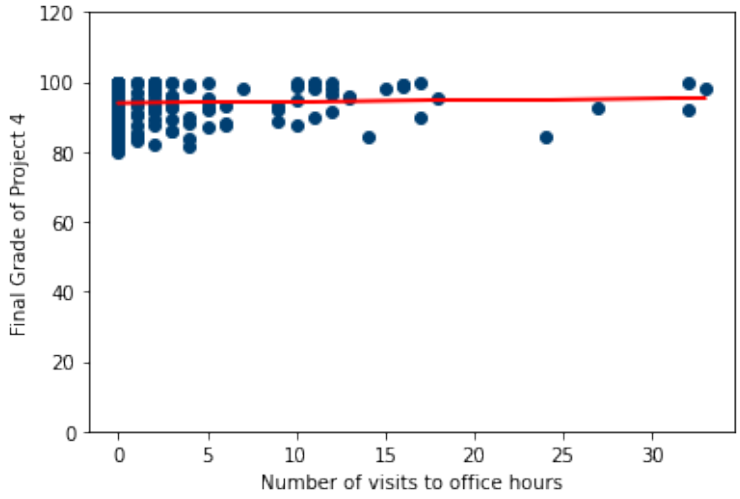

Figure 4.6: Non-struggling student data scatter plot between number of visits Piazza and final grade of project 4

the students came to receive direct help. 45.90% of students who struggled with Project 4 received direct instruction from instructors or TAs, while 52.7% of non-struggled students did. Among those who received help, the percentage of coming to the office in the first week of the project release date was 32.14% for struggling students and 33.3% for non-struggling students. The percentage of students who received direct help from instructional staff and students who started in the first week of the project release date were very similar. However, there were a few differences in patterns between struggling and non-struggling students in their management in receiving help. Struggling students tend to start coming into office hours earlier than non-struggling students by one day. For the rest of the differences, Table 4.10 shows the difference in each section of students' patterns.

We analyzed students who struggled and who did not struggle with Project 4. We compared the following categories between students who struggled with Project 4 or not; Project performance, Duration of office hour visits, and the number of visits. First, with the project performance, students who struggled had a mean score of 78.072 (with SD = 16.57), whereas,

	Project Performance		Duration of office hour visits (Min)		# of visits	
	Mean	SD	Mean	SD	Mean	SD
Struggled	78.07279484	16.57	38.4943	54.2252	3.97183	
Non-Struggled	94.04213929	5.54	49.7217	61.8831	2.2609	
T-test	13.3695		3.9825		-2.3965	
Sig. (p-value)	**<0.01**		**<0.01**		**0.01714**	

Table 4.10: Average calculated values to office hour visits and Paired Sample Test on struggling and non-struggling students

for students who did not struggle, the mean score was 94.02 (with SD = 5.54). Then we ran a T-test to see if the difference between the mean score value was valid, and we received T-test = 14.707 and P-value = $7.41E^{-34}$ which is less than 0.05. We do conclude that struggling students received a lower grade on Project 4 than non-struggled students, and this difference is significant. There was a difference in the duration of the office hour visits between the two groups of students. Students who we argue to have struggled with Project 4 have a mean duration of the office hour of 38.494 minutes (with SD = 54.22), whereas, for non-struggled students, the mean duration of office hour visits was 49.7216. As we ran the T-test, T-test = 3.983 and P-value = $4.59419E^{-31}$. As a result, we conclude that the duration of the office hour visits is shorter for struggling students. Lastly, on the mean number of office hour visits, struggled students had a mean number of office hour visits of 3.971 times (with SD =), and non-struggled students had a mean number of office hour visits of 2.2608 times (with SD =). The calculated T-test = -2.3965 and P-value = 0.017135169, which is also less than our set P-value. Therefore, we conclude that struggling students receive a lower average Project 4 score, have shorter office hour visits, and visit office hours more often than non-struggling students.

4.3 Piazza Data Analysis on Students who Struggled with Project 4

We used the same process on the Piazza data to find the relationship between usage of Piazza and Project 4 grade. We used original data of 488 students on how many times each student viewed Project 4 related questions on Piazza and Project 4 data to find the relationship between the two variables. Unfortunately, our first correlation value came out as 0.285, which means we conclude there is no relationship between the number of views on Project 4 related questions on Piazza and Project 4 grade. After sorting out struggling students, we created Spearman's correlation again, finding a higher correlation than using the original raw data. The Spearman's correlation value came out to be 0.388, and we can conclude that there is a slightly positive relationship between Project 4 question views on Piazza and academic outcome. Figure 4.7 is the scatter plot and the linear regression line using original 488 student data, and Figure 4.8 is the scatter plot and the linear regression using struggling student data. In contrast, Spearman's correlation value between non-struggling students' help-seeking behavior and their project performance came out as 0.142. Figure 4.9 represents the scatter plot with linear regression on students who did not struggle on Project 4. Calculated linear regression on Piazza data and it's significance test is shown in Table 4.11.

- Original Piazza data linear regression: $y = 0.122x + 84.805$
- Struggling student Piazza data linear regression: $y = 0.2193x + 73.8424$
- Non-struggling student Piazza data linear regression: $y = 0.02497x + 93.5789$

We then calculated linear regression analysis to find if the comparison between dataset is significant. Table 4.11 shows the result of the calculated linear regression analysis on the

	Original Data	Struggling student Data	Non-struggling student Data
Slope	0.122	0.21934367	0.024970
R-value	0.2171	0.3580025	0.10612
STD ERR	0.02489	0.136217	0.012715
P-value	$< \mathbf{0.01}$	$< \mathbf{0.01}$	0.0784

Table 4.11: Piazza Data Linear Regression Test

dataset. The calculated R-value for original data was 0.2193, whereas for struggling student data, it was 0.3565, and for the non-struggling student, it was 0.02497. P-value between original data, struggling student data, and non-struggling student data came out as $1.2459E^{-06}$, $7.7314E^{-08}$, 0.0784 in order. We can conclude that the difference between the relationship of the number of post views on Piazza and academic performance is moderately high between original data and struggling student data. We can conclude that the linear regression on the struggling student does show a positive relationship between Piazza interaction and academic performance on Project 4.

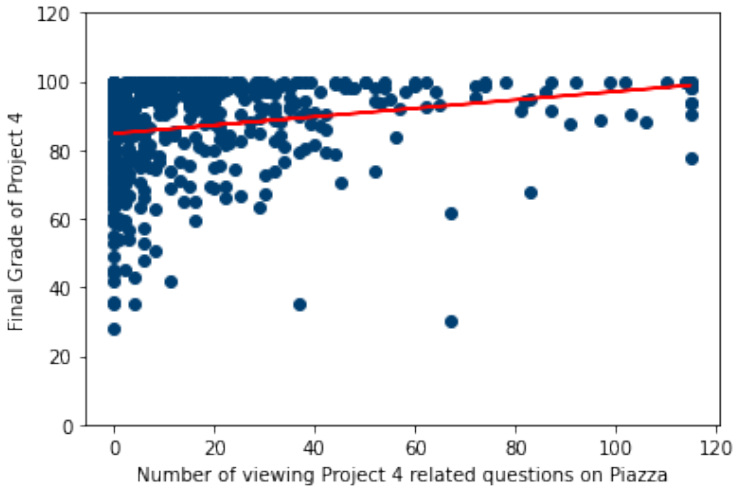

Figure 4.7: Original data scatter plot between number of visits Piazza and final grade of project 4

We also compared struggling and non-struggling student activities on Piazza to see if there were any differences in their activities. Table 4.12 shows the difference between two groups

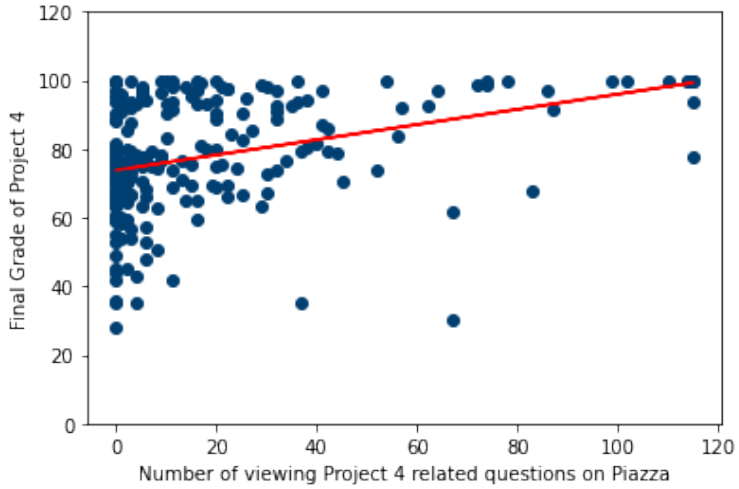

Figure 4.8: Struggling student data scatter plot between number of visits Piazza and final grade of project 4

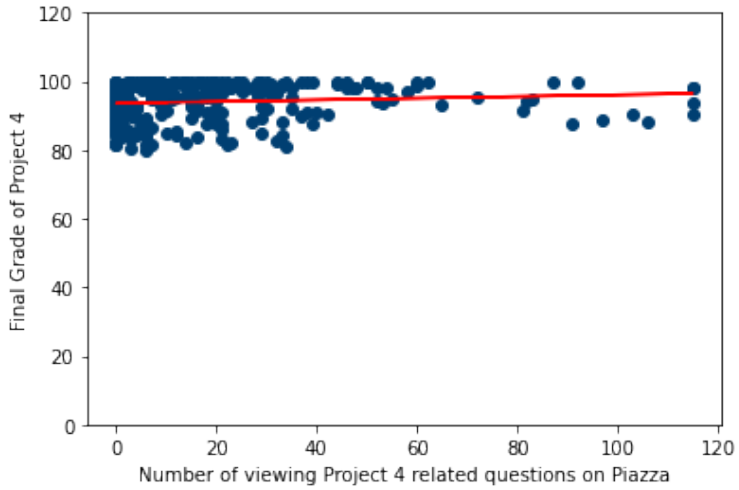

Figure 4.9: Struggling student data scatter plot between number of visits Piazza and final grade of project 4

of students' Piazza activities. Mean post views on Piazza Project 4 related questions of non-struggling students came out as 18.551 times (with SD 27.318) and for struggling students 18.8122 times (with SD 23.5758). However, the T-test between two students on the number of post views came out as -0.111 and P-value came out as 0.9114, which we do not consider as significant. For posting questions on Piazza, non-struggling students had slightly higher

	# of Post Views		# of Question posted	
	Mean	SD	Mean	SD
Non-Struggled students	18.55072464	27.318	0.2862	0.8074
Struggled students	18.8122	23.5758	0.2723	0.9232
T-test	-0.1113		0.1776	
Sig. (P-value)	0.9114		0.8591	

Table 4.12: Average number of Piazza activities and Paired Sample Test on Struggling students and Non-struggling students

number of average question posted on Piazza, yet the T-test value came out as 0.17766 and P-value as 0.85906, which we cannot argue that non-struggling students post more questions on Piazza. Figure 4.10 reflects the key findings from this section by HP and LP and struggling and non-struggling students.

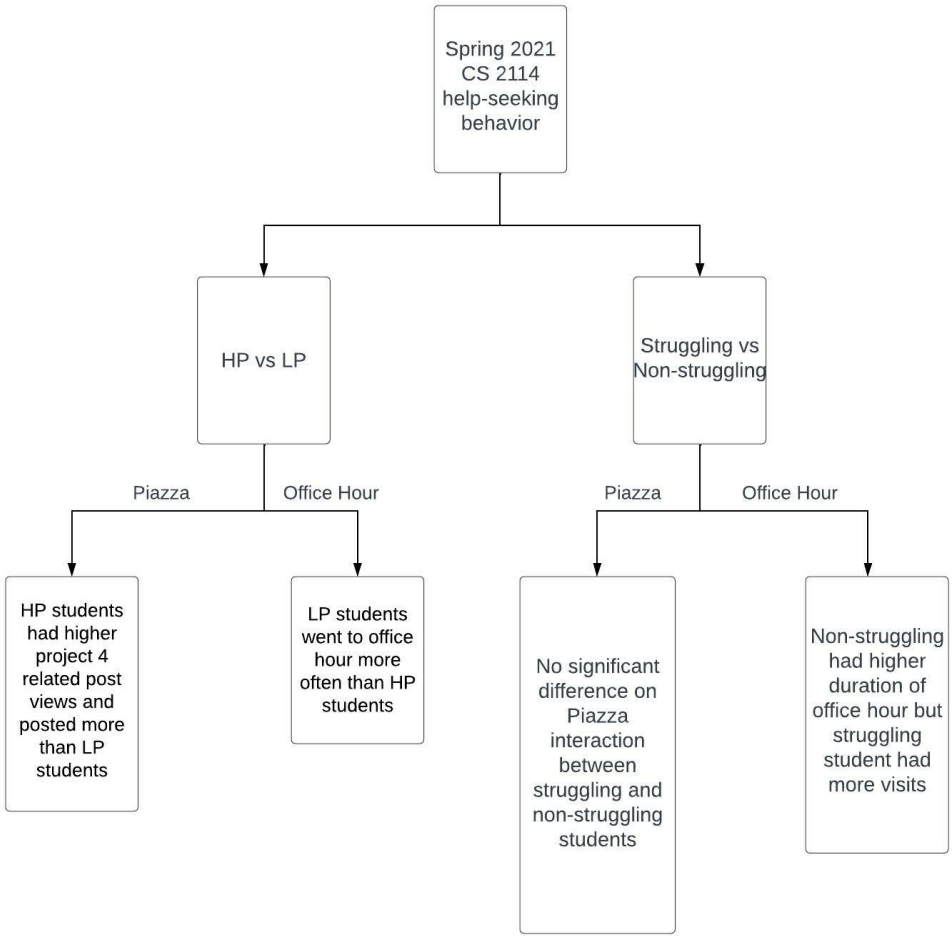

Figure 4.10: Key findings from Analyzing Help-seeking data

Chapter 5

Fall 2021 Survey data analysis

In this section we discuss the results of the Survey data taken on the Fall 2021 CS 2114 course. We were able to collect survey data from 347 students who agreed to participate. Figure 5.1 shows how difficult students perceived the project to be. Spring 2021 and Fall 2021 CS 2114 had same Project 4. 79.5% (276 students) either thought the project to be difficult or very difficult. We have done two analysis on different groups on the survey data. First analysis is on HP and LP students. HP students are students who self-reported of getting above 80 which we consider a good score, LP students are students who self-reported of getting below 80 which we believe a low performance. Second analysis is on two groups of students who perceived the project difficult or not. As the survey was taken anonymously, we could not apply the rules we used to find who struggled with the project. Therefore, we compared survey results by students' self-reported project score and self-reported project difficulty.

5.1 Survey Data analysis between HP and LP students

We divided students into two groups, HP and LP. Then we compared the number of students who perceived the project to be difficult. It is clearly shown that high-performing students have a higher percentage of thinking the project is not difficult or easy than the low-performing students. 99.1% of the students in LP students agreed that the project was

either difficult or very difficult, whereas the percentage of HP students finding the project difficult or very difficult was 72.8%. Table 5.1 shows the result from the survey on HP and LP students.

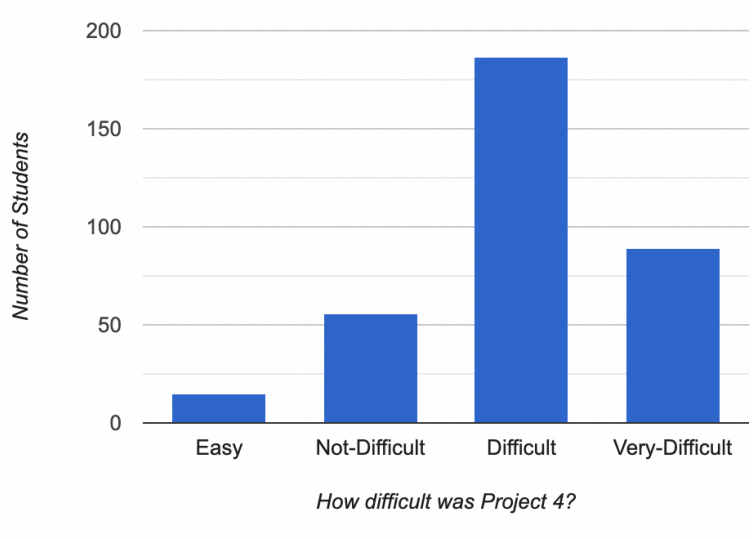

Figure 5.1: How difficult students perceived the project

We wanted to test for significant differences between categories, therefore we used cross tabulation. Table 5.2 represents the results between HP and LP students.

We compared different behaviors between HP and LP students. First, we compared lecture attendance between. Cross tabulation gave P-value = 0.0023, which we consider significant. 62% of HP students attended most or all lectures whereas 46% of LP students attended most or all lectures. For the number of times each group of students visited office hour for Project 4 between HP and LP, the P-value came out as 0.9484 which we do not consider significant. The difference on the number of office hour visits, 32% of HP students did not receive help whereas 34% of LP students did not receive help. Asked whether they received sufficient help from TAs, the P-value came out as less than 0.0001 and Chi-square 32.58, we consider the difference significant. The difference between HP and LP on if student received sufficient help from office hour, 89% of HP students agreed whereas 70% of LP students

Survey Question	Answer Choice	HP	%	LP	%
Lecture attendance	Only Few	30	13%	15	12%
	Less than Half	20	9%	21	17%
	More than Half	35	16%	30	25%
	Most Lectures	75	33%	40	33%
	All Lectures	65	29 %	16	13%
Project difficulty rate	Easy	14	6%	1	0.1%
	Not Difficult	49	22%	7	6%
	Difficult	129	57%	58	48%
	Very Difficult	33	15%	56	46%
Number of help received from TA	None	71	32%	42	34%
	0-5 times	101	45%	53	43%
	6-10 times	34	15%	18	15%
	More than 10 times	19	8%	9	7%
Did you receive sufficient help	Strongly Agree	71	32%	11	9%
	Agree	128	57%	75	61%
	Disagree	19	8%	30	25%
	Strongly Disagree	7	3%	6	5%
Staff help helped	Strongly Agree	35	16%	6	0.5%
	Agree	146	65%	62	51%
	Disagree	31	14%	42	34%
	Strongly Disagree	13	6%	12	10%
Total		225		122	

Table 5.1: Fall 2022 CS2114 Survey Results

agreed. For the question about whether the staff was helpful, the result came out as P-value < 0.0001 and Chi-square $= 28.03$ which we also consider significant. The difference on whether interaction with TA helped, 81% of HP students agreed where 52% LP students agreed. As a conclusion, we could conclude that HP students have higher rate in attending lecture, higher rate in how sufficient office hour was and higher rate in perceiving TAs were helpful than LP students.

Question	Sig. (P-value)	Chi-square	Carmer's V
Project score	**< 0.0001**	**49.45**	**0.3775**
Lecture	**0.0023**	**16.59**	**0.2187**
Office hour and piazza	0.9484	0.36	0.0322
sufficient help	**< 0.0001**	**32.58**	**0.3064**
interaction helped	**< 0.0001**	**28.03**	**0.2842**

Table 5.2: Cross tabulation with survey questions on HP and LP students

5.2 Survey Data analysis between students who perceived Project 4 difficult and not difficult

We also ran analysis between students who perceived Project 4 difficult or not to see any difference in their behavior. Table 5.3 shows the raw data for the two groups. Table 5.4 shows the results of running cross tabulation between students who found the project difficult and not difficult.

Students who perceived Project 4 difficult and not difficult had similar percentage in lecture attendance. For students who thought Project 4 was difficult, the percentage of attending most to all lectures was 56% whereas for considering project not difficult, the percentage of attending most to all lectures was 59%. The P-value between two groups in lecture attendance came out as 0.0008 and Chi-square = 40.71 which we consider the difference significant. However, the difference on lecture attendance was only 3%. For how many times did each student sought help from office hour on Project 4, 28% of students who perceived project difficult sought no help whereas, 58% students who did not perceive project difficult sought no help. The result on difference between two groups came out as P-value < 0.0001 and Chi-square = 37.15. Asking whether the student had sufficient help from office hour, 72% of students who found the project difficult agreed, whereas 94% of students who found

Survey Question	Answer Choice	Difficult	%	Not Difficult	%
Lecture attendance	Only Few	27	10%	18	25%
	Less than Half	35	13%	6	8%
	More than Half	60	22%	5	7%
	Most Lectures	93	34%	22	31%
	All Lectures	60	22%	20	28%
Project score	Below 60	53	19%	5	7%
	60 - 69	21	8%	0	0%
	70 - 79	40	15%	3	4%
	80 - 89	56	20%	6	8%
	90 and above	105	55%	57	80%
Number of help received from TA	None	77	28%	41	58%
	0-5 times	130	47%	24	34%
	6-10 times	47	17%	5	7%
	More than 10 times	27	10%	1	1%
Did you receive sufficient help	Strongly Agree	53	13%	28	39%
	Agree	163	59%	39	55%
	Disagree	47	17%	2	3%
	Strongly Disagree	11	4%	2	3%
Staff help helped	Strongly Agree	28	10%	5	7%
	Agree	171	62%	37	52%
	Disagree	58	21%	15	21%
	Strongly Disagree	20	7%	14	20%
Total		275		71	

Table 5.3: Fall 2022 CS2114 Survey Results between students who found Project 4 difficult and not difficult

project not difficult agreed. The result came out as P-value = 0.0004 and Chi-square = 18.27. Asked whether if the interaction with the TA helped, 72% of students who found the project difficult agreed, whereas, 59% of students who found project not difficult agreed. However, the result in P-value came out as 0.9078 and Chi-square 0.55 which we do not consider significant. Therefore, we could conclude that there was no difference on lecture attendance between students who found project difficult and not difficult, students who found project difficult had higher percentage in visiting office hours, and had higher percentage in agreeing they had sufficient help from TAs.

Question	Sig. (P-value)	Chi-square	Carmer's V
Project score	**< 0.0001**	**40.712**	**0.2885**
Lecture	**0.0008**	**19.02**	**0.2345**
Office hour and piazza	**< 0.0001**	**37.15**	**0.3204**
sufficient help	**0.0004**	**18.27**	**0.2301**
interaction helped	0.9078	0.55	0.0403

Table 5.4: Cross tabulation with survey questions on students who found project difficult and not difficult

Topic	LP	HP
Time management	4	4
Office hours(TAs) were helpful	2	13
Office hours(TAs) were not helpful	11	4
readiness related topic	1	1
Never went to office hours	0	4
Piazza was helpful	1	0
Project 4 feedback	0	2
Web-CAT feedback	0	1
None	4	4
Total	20	33

Table 5.5: Distribution on extra feedback Low Performers and High Performers gave

5.3 Survey Data analysis on feedback

Feedback variation between HP and LP students was also high. 53 out of 347 students decided to write feedback on the project. 33 students received higher than 80, while 20 students received less than 80 on the project. Table 5.5 shows the feedback from HP and LP students. Where HP students' feedback had 9 topics, LP students' feedback had 6 topics. HP students' feedback topics focused on office hour help and project, whereas LP students' feedback topics focused more on office hours. While the topic of feedback is similar, the core difference was in how each group of students reacted. While 40% of HP students said office hours were helpful, 10% of LP students mentioned that office hours were not helpful. In contrast, 12.5% of HP students and 55% of LP students agreed that office hour was unhelpful.

On the time management feedback, 12.5% of HP and 20% of LP students agreed on how each student procrastinated and should have managed time better. We could conclude that HP students find office hours more helpful than LP students.

Chapter 6

Future Work

Help-seeking behavior is an important activity for students. It is a regulatory skill for students to build good learning skills. We devised a procedure to distinguish students who display evidence that they struggled with a programming and we have found that high-performing students and non-struggling students do engage in more help-seeking activities. There were some limitations in this research. The first was that the survey had to be anonymous, and we could not verify which management skills each student used. The feedback topics varied from HP and LP students, and in the future, we can do various interviews on how each student managed the project and verify which skills are distributed. Regarding office hours, the interview can be specified on office hours on why HP students had a higher percentage of perceiving office hours helpful than LP students. Another limitation was Zoom-based log data. In the Zoom-based log data, we were able to find that the average duration of the office hour visits for HP and LP students varied. Yet, with just interaction data, we could not find the reason for the difference. In further research, we could interview students to ask about the reasons for the office hour visits. With the information on the reason for coming into office hours, we will be able to understand the differences in help-seeking activities for HP and LP students.

A collection of useful academic skills and strategies can be identified as the followings [7]:

- Time management
- Incremental development
- Self-checking (meta-cognition)
- Persistence
- Planning

Our research is focused on self-checking skills. we found a positive relationship between help-seeking behavior and academic performance in students who struggled with a project. Our research can be the basis for finding the relationship between other skill sets and academic performance. Even in self-checking, there are several skills that students can utilize. One of the skills that students can manage is help-seeking. We hope that our research becomes the basis for finding other relationships between other skill sets and academic performance to lead CS undergraduate students to success in academic performance.

Chapter 7

Conclusion

Help-seeking is a learner's meta-cognitive awareness in receiving help to satisfy their need to learn a new material [23]. Here we set out to develop insight into how the help-seeking behaviors affect students in the CS2 course. Analyzing students' data from CS2 courses has confirmed that HP students engage in more effective help-seeking behavior than LP students. Our research studied students' interaction data of help-seeking behaviors, such as receiving help from instructional sources like online forums, instructional staff, and search engines. By using 488 student data from Spring 2021 CS 2114 course and 347 student survey data from Fall 2021 CS 2114 course data, we answered the following research question addressed in Chapter 1. Comparing raw student interaction data of help-seeking behavior on office hour visits and Piazza usage, we were unable to find any relationship between help-seeking behavior and academic outcome. However, when focusing on finding students who struggled with Project 4, we were able to find the relationship between help-seeking behaviors and academic outcomes.

To find the relationship between help-seeking behaviors and academic performance, we first had to collect necessary student interaction data. We decided to use Project 4 data in the 2021 Spring and Fall CS 2114 course because it is a second-year CS course required for CS students that teaches fundamental data structure concepts. Project 4 is the most challenging individual project in the course. In Spring 2021 CS 2114, we collected help-seeking behaviors and Project 4 academic performance. The help-seeking behavior includes office hour visits

and Piazza usage. For Fall 2021, we ran survey questions related to Project 4 shown in Chapter 3 survey data collection. We analyzed two different semesters of data separately. First, we worked to find the relationship between help-seeking activities and then analyzed the survey. When we first created Spearman's correlation between help-seeking activities and Project 4 grade, we could not find any relationship. After a thorough discussion, we found that two groups of students who struggled with the project could benefit from using help-seeking activities and who did not struggle with Project 4. After the discussion, we found ways to sort out who might have struggled with Project 4, and we argue they are the students who will benefit from using office hours. We divided students into two groups, those who received higher than 80 as HP and less than 80 as LP. We first categorized the LP students as the students who needed help, but we needed to find students who also perceived the project to be difficult within the HP group. Here we deployed our algorithm of analysis on student behaviors. We applied the rule to all HP students to find struggling students. The rules were if a student's submission is over 10 times and the project correctness lower than 50% by the third submission. We found 213 students who we believe in having struggled with Project 4.

We also analyzed the results of our survey data. We used survey data on how each student perceived the project as difficult and compared the percentage of students who we sorted out to have struggled and the percentage of students who said Project 4 was difficult. With the students who we believe in having struggled with Project 4, we created Spearman's correlation once again, which came out as follows in Table 7.2. Spearman's correlation between help-seeking behaviors on students who did not struggle is shown in Table 7.3. Survey data was also used in analyzing the help-seeking behavior of HP and LP students. We could see the difference in help-seeking patterns in HP and LP students. The findings on different patterns in help-seeking behavior on a different group of students are as following:

- High Performing students seek more help than Low Performing students in Piazza but for the office hour it is similar between HP and LP
- Struggling students seek more help from Office Hours while non-struggling students seek more help from Piazza
- Students who found project difficult seek more help than students who do not
- Help-seeking behavior has positive relationship with academic performance on struggling students

To address research question 1 (RQ1), a strong relationship exists between help-seeking activities and academic performance in students who struggled with an assignment. While we could not find the relationship between help-seeking behavior and academic performance in LP students, we could find students who struggled with Project 4 had a positive relationship between help-seeking activities during office hours and usage of the online forum Piazza and academic performance. We created Spearman's correlation to prove the relationship between two variables. The correlation between help-seeking activities and academic outcomes in original data is shown in Table 7.1 in students who struggled with Project 4 is shown in Table 7.2, and for students who did not struggle with Project 4, the correlation is displayed in Table 7.3.

	Spearman's correlation value	Sig. (P-value)
# of Office hour visits	0.079	0.08108
# Project 4 post views on Piazza	**0.285**	**< 0.01**

Table 7.1: Calculated Spearmans' correlation between two help-seeking behaviors and Project 4 grade on original data

We could conclude that students who struggled with Project 4 positively correlated with help-seeking activities and Project 4 grades. In contrast, those who did not struggle with

	Spearman's correlation value	Sig. (P-value)
# of Office hour visits	**0.313**	**< 0.01**
# Project 4 post views on Piazza	**0.412**	**< 0.01**

Table 7.2: Calculated Spearmans' correlation between two help-seeking behaviors and Project 4 grade on students who struggled with Project 4

	Spearman's correlation value	Sig. (P-value)
# of Office hour visits	-0.022	0.5155
# Project 4 post views on Piazza	0.165	0.0784

Table 7.3: Calculated Spearmans' correlation between two help-seeking behaviors and Project 4 grade on students who did not struggle with Project 4

Project 4 did not have any relationship between help-seeking activities and Project 4 grade. We argue that students who struggled with Project 4 would benefit from utilizing help-seeking behaviors.

For research question 2 (RQ2) and research question 3 (RQ3), there is a large difference in help-seeking activities between HP and LP students and between students who struggled with the project and students who did not. By observation, HP students sought more help than LP students in Piazza, while LP students sought more help during office hours than HP students.

In research question 4 (RQ4), we could define struggling students by using the algorithm shown in Figure 4.3. We concluded that struggling students' help-seeking behavior positively correlates with academic performance.

Lastly, in research question 5 (RQ5), by comparing each group of students' help-seeking activities and survey data, we could find that students who struggled with Project 4 and students who did not struggle with Project 4 also showed a difference in help-seeking activities. Students who struggled sought more help from both office hours and Piazza than students who did not struggle. By answering these research questions, we have concluded that help-seeking behavior like receiving help during office hours and Piazza on students

who have struggled with a project has a positive relationship with academic performance. Struggling students can benefit from utilizing the help resources provided. Another factor we could conclude with our research was that HP students engage more in help-seeking than LP students do.

Bibliography

[1] T. Bartholomé, E. Stahl, S. Pieschl, and R. Bromme, "What matters in help-seeking? a study of help effectiveness and learner-related factors," *Computers in Human Behavior*, vol. 22, pp. 113–129, 01 2006.

[2] P. Biró and M. Csernoch, "Deep and surface metacognitive processes in non-traditional programming tasks," in *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*, 2014, pp. 49–54.

[3] R. Deloatch, B. P. Bailey, A. Kirlik, and C. Zilles, "I need your encouragement! requesting supportive comments on social media reduces test anxiety," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 736–747. [Online]. Available: https://doi.org/10.1145/3025453.3025709

[4] A. Doebling and A. M. Kazerouni, "Patterns of academic help-seeking in undergraduate computing students," in *21st Koli Calling International Conference on Computing Education Research*, ser. Koli Calling '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3488042.3488052

[5] A. DONG, M. Jong, and R. King, "How does prior knowledge influence learning engagement? the mediating roles of cognitive load and help-seeking," *Frontiers in Psychology*, vol. 11, p. 591203, 10 2020.

[6] S. H. Edwards, "Automated feedback, the next generation: Designing learning experiences," in *Proceedings of the 52nd ACM Technical Symposium on Computer*

Science Education, ser. SIGCSE '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 610–611. [Online]. Available: https://doi.org/10.1145/3408877.3437225

[7] S. H. Edwards and Z. Li, "A proposal to use gamification systematically to nudge students toward productive behaviors," in *Koli Calling '20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*, ser. Koli Calling '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3428029.3428057

[8] D. Ginat, "Metacognitive awareness utilized for learning control elements in algorithmic problem solving," in *Proceedings of the 6th Annual Conference on Innovation and Technology in Computer Science Education*, ser. ITiCSE '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 81–84. [Online]. Available: https://doi.org/10.1145/377435.377490

[9] Q. Hao, B. Barnes, R. Branch, and E. Wright, "Predicting computer science students' online help-seeking tendencies," *Knowledge Management & E-Learning: An International Journal*, vol. 9, pp. 19–32, 03 2017.

[10] Q. Hao, E. Wright, B. Barnes, and R. Branch, "What are the most important predictors of computer science students' online help-seeking behaviors?" *Computers in Human Behavior*, vol. 62, pp. 467–474, 09 2016.

[11] P. Kinnunen and B. Simon, "Cs majors' self-efficacy perceptions in cs1: Results in light of social cognitive theory," in *Proceedings of the Seventh International Workshop on Computing Education Research*, ser. ICER '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 19–26. [Online]. Available: https://doi.org/10.1145/2016911.2016917

[12] A. Kitsantas and A. Chow, "College students' perceived threat and preference for seeking help in traditional, distributed, and distance learning environments," *Computers & Education*, vol. 48, no. 3, pp. 383–395, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131505000199

[13] S. N. Liao, S. Valstar, K. Thai, C. Alvarado, D. Zingaro, W. G. Griswold, and L. Porter, "Behaviors of higher and lower performing students in cs1," ser. ITiCSE '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 196–202. [Online]. Available: https://doi.org/10.1145/3304221.3319740

[14] D. Loksa and A. J. Ko, "The role of self-regulation in programming problem solving process and success," in *Proceedings of the 2016 ACM Conference on International Computing Education Research*, ser. ICER '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 83–91. [Online]. Available: https://doi.org/10.1145/2960310.2960334

[15] D. Loksa, A. J. Ko, W. Jernigan, A. Oleson, C. J. Mendez, and M. M. Burnett, "Programming, problem solving, and self-awareness: Effects of explicit guidance," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1449–1461. [Online]. Available: https://doi.org/10.1145/2858036.2858252

[16] A. Maier, C. Schaitz, J. Kroener, A. Berger, F. Keller, P. Beschoner, B. Connemann, and Z. Sosic-Vasic, "The association between test anxiety, self-efficacy, and mental images among university students: Results from an online survey," *Frontiers in Psychiatry*, vol. 12, 11 2021.

[17] D. J. Malan, "Virtualizing office hours in cs 50," in *Proceedings of the 14th Annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education*,

ser. ITiCSE '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 303–307. [Online]. Available: https://doi.org/10.1145/1562877.1562969

[18] S. Marwan, A. Dombe, and T. W. Price, "Unproductive help-seeking in programming: What it is and how to address it," in *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, ser. ITiCSE '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 54–60. [Online]. Available: https://doi.org/10.1145/3341525.3387394

[19] C. Midgley, A. Ryan, and P. Pintrich, "Avoiding seeking help in the classroom: Who and why?" *Educational Psychology Review*, vol. 13, 06 2001.

[20] P. Moliterni, S. DE STASIO, M. Carboni, and D. Carlo, "Motivational and self regulated learning components of academic performance," 05 2010.

[21] J. Prather, R. Pettit, K. McMurry, A. Peters, J. Homer, and M. Cohen, "Metacognitive difficulties faced by novice programmers in automated assessment tools," in *Proceedings of the 2018 ACM Conference on International Computing Education Research*, ser. ICER '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 41–50. [Online]. Available: https://doi.org/10.1145/3230977.3230981

[22] T. W. Price, Z. Liu, V. Catet, and T. Barnes, "Factors influencing students' help-seeking behavior while programming with human and computer tutors," ser. ICER '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 127–135. [Online]. Available: https://doi.org/10.1145/3105726.3106179

[23] P. Sands and A. Yadav, *Self-Regulation for High School Learners in a MOOC Computer Science Course*. New York, NY, USA: Association for Computing Machinery, 2020, p. 845–851. [Online]. Available: https://doi.org/10.1145/3328778.3366818

[24] C. A. Schmidt-Jones, "An open education resource supports a diversity of inquiry-based learning," *The International Review of Research in Open and Distributed Learning*, vol. 13, no. 1, pp. 1–16, Jan. 2012. [Online]. Available: http://www.irrodl.org/index.php/irrodl/article/view/1141

[25] M. Smith, Y. Chen, R. Berndtson, K. Burson, and W. Griffin, ""office hours are kind of weird": Reclaiming a resource to foster student-faculty interaction," *InSight: A Journal of Scholarly Teaching*, vol. 12, pp. 14–29, 08 2017.

[26] B. E. Vaessen, F. J. Prins, and J. Jeuring, "University students' achievement goals and help-seeking strategies in an intelligent tutoring system," *Comput. Educ.*, vol. 72, no. C, p. 196–208, mar 2014.

[27] Y. Yang and J. Taylor, "The role of achievement goals in online test anxiety and help-seeking," *Educational Research and Evaluation*, vol. 19, no. 8, pp. 651–664, 2013. [Online]. Available: https://doi.org/10.1080/13803611.2013.811086

[28] B. J. Zimmerman, "Becoming a self-regulated learner: Which are the key subprocesses?" *Contemporary Educational Psychology*, vol. 11, no. 4, pp. 307–313, 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0361476X86900275