Let's Talk It Out: A Chatbot for Effective Study Habit Behavioral Change

XIAOYI TIAN, University of Florida, USA
ZAK RISHA, University of Pittsburgh, USA
ISHRAT AHMED, University of Pittsburgh, USA
ARUN BALAJIEE LEKSHMI NARAYANAN, University of Pittsburgh, USA
JACOB BIEHL, University of Pittsburgh, USA

Research has shown study habits and skills to be correlated with academic success, calling for a deeper comprehension of these behaviors and processes to design effective interventions for struggling students. Chatbots have recently been used as a persuasive technology to help support behavioral change, making them an intriguing design space for students' study habits and skills. This paper investigated the feasibility of using chatbots for promoting behavioral change of college students majoring in Computer Science (CS). We conducted semi-structured interviews with CS peer-tutors and surveyed university freshmen to understand students' study habits and identify technical intervention opportunities. Inspired by the findings, we designed *StudyBuddy*, a chatbot prototype deployed in Slack that periodically sends tips, provides assessments of students' study habits via surveys, helps the students break down assignments, recommends academic resources, and sends reminders. We evaluated the usability of the prototype in-depth with 8 students (both first-year and senior students) and 5 course instructors followed by a large scale evaluative survey (n=117) using video of the prototype. Our research identified important design challenges such as building trust and preserving privacy, limiting interaction costs, and supporting both immediate and long-term sustainable support. Likewise, we proposed design recommendations that demonstrate context awareness, personalize the experience based on user preferences, and adapt over time as students mature and grow.

CCS Concepts: • **Human-centered computing → Empirical studies in interaction design**.

Additional Key Words and Phrases: chatbot; study habits and skills; behavioral change; persuasive technology; computer science

ACM Reference Format:
Xiaoyi Tian, Zak Risha, Ishrat Ahmed, Arun Balajiee Lekshmi Narayanan, and Jacob Biehl. 2021. Let's Talk It Out: A Chatbot for Effective Study Habit Behavioral Change. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 97 (April 2021), 32 pages. https://doi.org/10.1145/3449171

1 INTRODUCTION

According to the Institute of Education Sciences, the graduation rate of college students over the past 6 years in the US is only 60% [45]. These high attrition rates are troubling and pose problems for filling high demand labor markets like STEM. The U.S. National Science and Technology Council's

Authors' addresses: Xiaoyi Tian, University of Florida, 32611, USA, Gainesville, FL, tianx@ufl.edu; Zak Risha, University of Pittsburgh, 15213, USA, Pittsburgh, PA, zak.risha@pitt.edu; Ishrat Ahmed, University of Pittsburgh, 15213, USA, Pittsburgh, PA, isa14@pitt.edu; Arun Balajiee Lekshmi Narayanan, University of Pittsburgh, 15213, USA, Pittsburgh, PA, arl122@pitt.edu; Jacob Biehl, University of Pittsburgh, 15213, USA, Pittsburgh, PA, biehl@pitt.edu.

2573-0142/2021/4-ART97 $15.00
https://doi.org/10.1145/3449171

Committee on STEM Education issued a 2018 report noting that while demand for STEM graduates continues to grow, other countries like India and China are doing a better job preparing students and now produce nearly half of all STEM graduates [14]. Transition to university comes with many potential life changes such as relocation, long-distance relationships, greater responsibility and independence, and increased academic rigor [9]. Navigating through the new environment with these additional challenges can be frustrating, and students may lack the necessary skills to survive in the university's competitive environment. Similarly, inadequate study habits and skills, lack of mentoring, low motivation, and low self-efficacy can impair success in the more independent university environment [7]. These deficiencies may cause students to lose interest in their coursework and abandon their studies, resulting in higher attrition rates. While not new, these challenges, remain a compelling problem, one that we believe can be addressed by the emergence of new technologies.

While evidence of high attrition rates points to inadequate preparation for university studies, psychological and behavioral components, i.e., study habits and skills, are also important to consider. Study habits and skills are a group of constructs often studied in relation to academic performance [7]. An important distinction between study behavior and study habits is that the latter is repeated regularly and takes time to instill. Existing literature provides evidence that study habits and skills are correlated with academic success [15, 38]. Self-regulated learning, which includes study habits and skills, also has been modeled as a key component to academic achievement [46, 64]. Students with poor study habits and skills manage time poorly, engage in ineffective learning strategies, maintain unhealthy routines, or fail to reflect on their learning processes [46]. All of these traits can influence academic performance as well as lead to psychological repercussions on anxiety level, attitude, self-efficacy, and motivation. In particular, University freshmen transition into a new environment that requires more independent academic work, making these study habits and skills potentially a critical factor. We believe technology can provide beneficial support to these students, aiding the development of effective study habits and skills.

In this paper, we aim to investigate the use of emerging technology to improve the students' study behaviors and skills that could lead to positive habit formation. Universities are an inherently cooperative and collaborative environment that presents opportunities for technology to help support students' engagement in their work. While educational technology has had success in improving academic performance, psychological and behavior improvements are still an emerging research area. Du Boulay et al. [18] highlighted the importance of going beyond cognition, taking into account other factors such as motivation, metacognition, and emotion to create "educational systems that care." Motivated by this work, we explore the design and use of chatbots for influencing the metacognition of first-year university students majoring in Computer Science (CS). We are particularly interested in chatbots as they already have been used to alter user behavior and foster self-awareness [25, 28, 49]. Given the remote learning environments under the COVID-19 pandemic, students require more support with grasping knowledge effectively, managing time, keeping track of assignments, and staying connected to their learning community, thus giving chatbots greater potential to facilitate learning and behavioral change in a remote setting.

We surveyed students and interviewed tutors to better understand students' issues related to study habits and skills and identified technological intervention opportunities. These findings informed several novel design features for chatbot-based study behavior tools and inspired us to create *StudyBuddy*, a chatbot prototype deployed on Slack to influence students' study behaviors. The functionality of *StudyBuddy* includes sending tips from experienced students, providing assessment and feedback of students' study habits via surveys, helping the students break down academic assignments, recommending academic resources, and sending reminders to complete assignments on time. We evaluated *StudyBuddy* using in-depth student and instructor interviews, and student surveys, contrasting the feedback we received from both first-year and upper-class students, and instructors. The results identified important design challenges such as building trust and preserving privacy, limiting interaction costs, and supporting both immediate and long-term sustainable support. Likewise, our findings help generate design recommendations for a chatbot that demonstrates context awareness, personalizes the experience based on user preferences, and adapts over time as students mature and grow.

2 RELATED WORK

2.1 Persuasive Technology for Behavioral Change

Prior research has sought to develop technology to alter everyday behavior, constituting a significant body of literature known as persuasive technology [23]. The task of getting an individual to change everyday behaviors, however, is difficult, which is why research has focused on the design of various technologies [13, 17, 24, 33, 53]. A common application of persuasive technology has been promoting behavior related to health and wellness, which encompasses a large portion of the literature [13, 44, 52, 53]. Despite the growing popularity of persuasive technology, use and adoption of these technologies has presented many challenges. In a six-week study where college students were asked to use Fitbit activity trackers, participants abandoned technology quickly, with over 50% drop out rate in two weeks [43]. Researchers investigated why users abandon certain smart devices and found several factors responsible: devices not aligning with users' conceptions of themselves, data collected perceived not to be useful, and device maintenance becoming unmanageable [39]. There was also another type of "happy abandonment" in health tracking devices where individuals successfully achieve their goals, thus do not need their devices anymore [10]. However, research applying persuasive technology to behaviors related to education, like study habits, is much more scarce. Previous designs and methods applied in a different domain might not be as applicable in education, especially computing education, calling for more research on how to design these technologies and understand the benefits and the challenges for students.

2.2 Education Interventions Supporting Computer Science or Study Habits

In education, many researchers and educators have contributed tools or advanced techniques to address needs in computing education. For example, Online Python Tutor [26] and PRIME [20] are entry-level programming tools designed for promoting computational thinking and providing adequate feedback. Another popular way for CS students to obtain help on syntactic and conceptual knowledge by watching online lecture videos from platforms like YouTube and MOOCs [34]. In addition, online Q&A communities like StackOverflow[1] or discussion forums such as Piazza, allow students to learn more about programming while helping others [50, 56]. However, these tools and techniques tend to focus on supporting domain knowledge, which is not always helpful in solving students' problems if they lack adequate programming strategies (e.g., testing strategies) [54] and effective study habits. Thus, we identify a need to explore how to design technology that helps students form effective study skills and behaviors that facilitate learning computer science's strategic and behavioral aspects .

While limited, there has been prior research designing and developing interventions to improve study habits. Theoretically, Filippou attempted to apply the *Fogg Behavioral Model*, a model for persuasive design, to an application that promotes better study habits [22]. Their design, centered around study scheduling, class preparation, and group study, was inspired by theoretical frameworks, models, and educational literature. However, they had yet implemented their design or conducted any user testing. Practically, Kreynin et al. [36] developed an SMS chatbot that enables

[1]http://stackoverflow.com

reflective journaling for first-year engineering students to support positive habit formation. While an SMS platform is suitable for journaling, the platform offers fewer design affordances than other technologies, limiting the types of features and interactions possible. For instance, other chat applications, like Slack, allow for interactive messages that incorporate user interface elements, such as buttons, modals, menus, and date pickers, to enrich the capabilities of the chat. This potential enrichment calls for more research investigating a chatbot that incorporates a wider array of interactive features.

2.3 Chatbot Applications for Behavioral Intervention

Chatbots have also been used as a persuasive technology to influence the daily behaviors of users in other contexts including improving sleep habits [55], providing psychiatric counseling [51], promoting diet and exercise [29], helping to stop smoking [5], and other health-related interventions [1, 31]. In CSCW, Lukoff et al. utilized a chat-based food journaling tool for families to support their healthy eating goals [42]. Additionally, efforts have been made to incorporate chatbots into the workplace by helping users reflect and learn from their experiences [35] or helping users psychologically detach and reattach from work, facilitating recovery and well-being [63]. This growing literature suggests chatbots may be equally suitable to the educational domain. Not surprisingly, researchers have designed pedagogically focused chatbots that provide individual support to student forum posts [25] or recommend subjects to users in MOOCs [28, 49]. In terms of psychologically focused agents, Zvereva et al. [65] proposed a dialogue-based method to assess student motivation and provide necessary recommendations to improve a student's motivation. Despite this recent work, designing a chatbot to influence study habits and skills remains unexplored. This leads to a potentially significant research opportunity given the popularity of using chatbots for behavioral intervention and the key role study habits and skills play in academic performance.

3 DESIGN INQUIRY

Our approach to inform our design was both user and domain-centric, focusing on identifying behaviors detrimental to academic performance and factors critical to student success. We first conducted preliminary semi-structured interviews with CS peer-tutors. The findings from these interviews informed a subsequent survey that helped understand students' study behaviors and desirable functionality of the technical interventions. We sought to formalize the different features that we would like to incorporate in our final design of the chatbot through design inquiry. The flow of the entire design and evaluation process is demonstrated in the *Appendix*. All studies in this paper were conducted at the Department of Computer Science of a public research university located in the Northeastern United States.

3.1 Preliminary Interview

We conducted semi-structured interviews with three peer tutors. These peer tutors were junior or senior undergraduate students with strong academic records employed in the CS department tutoring center. They primarily help freshmen or sophomore students with their course projects or assignments. The peer tutors we interviewed closely interacted and observed students over a longer period of time (at least one semester). The tutors were recruited by direct emails and no compensation was offered. The interview lasted about 20 minutes. One of the researchers typed notes during the interview. Our interview questions were guided with the following broad-level questions in mind:

- What are the different types of students coming into the peer tutoring system?
- What factors motivate students to seek help at the peer tutoring center?

- How do peer tutoring sessions impact students academically and behaviorally?
- What are the observed challenges and shortcomings students have in regards to study habits and skills?

Through these interviews, we gained useful insights into freshman study habits and domain-specific problems. The researchers qualitatively analyzed the interview notes to extract themes to address the above questions. Peer-tutors classified the students by the frequency of attendance into two categories: repeated tutees and occasional tutees. Repeated tutees regularly attended tutoring sessions with clear goals and demonstrated time management. Peer tutors were able to observe gradual progress being made as well an academic improvement for the repeated tutees. In contrast, occasional tutees were frequently lost during their assignments and needed immediate support to complete their assignments. Such observed differences also relate to other questions mentioned above, e.g., tutors noted that the assignment deadlines motivated students to come to the peer tutoring center and procrastination was a common shortcoming in students' behavior and habits. Other factors that motivated students' attendance in the peer tutoring center included preferring one-to-one interaction, asking specific questions, or facing difficulty with starting an assignment. Tutors also observed specific domain related deficiencies e.g., not being able to compile code. Similarly, many students just wanted help debugging the code, suggesting they lack certain domain knowledge to individually approach this task. In addition to these issues, tutees also would express a lack of confidence with being successful at the coursework.

We synthesized the findings from these semi-structured interviews into three common challenges: **time management**, **task-goal management** and **the lack of domain knowledge**. We identified these shortcomings as useful targets for intervention to help students develop autonomy and self-efficacy in handling the academic challenges faced in their line of study. Following the evidence from these tutor interviews, the research team designed solutions to address the peer tutors' concerns, that is, improving task and time management, providing peer feedback, and recommending appropriate resources.

3.2 Survey Design and Distribution

Beyond informing our proposed design features, our findings from the peer interview helped us develop a survey for first-year CS students to gain insight into their study behaviors. More specifically, the survey asked about students' time and goal management, self-efficacy for CS courses, and strategies using academic resources. We also inquired about the potential of a chatbot as a form of intervention, as well as their perceptions of our proposed design solutions. Our survey questions, included in the *Appendix*, consisted of questions regarding usage of messaging apps, including Slack [2] as it is a notable team collaboration platform adopted widely in academic environments. Our self-efficacy questions were adapted from MSLQ [19]. The remaining questions were constructed on a 5-point Likert scale to assess study behaviors, familiarity with academic resources, and usefulness of different chatbot design features. Surveys were distributed in an introductory computer science course at the same university over two consecutive semesters (Fall 2019 and Spring 2020). With the course instructor's permission, the researcher came into the classroom and shared the survey link to solicit voluntary responses. The average completion time of this short survey was 3 minutes.

3.3 Survey Findings

A total of n=83 valid responses were collected, 61 were collected in Fall 2019 (the 1st semester), and 22 were collected in Spring 2020 (the 2nd semester). A two-tailed t-test revealed no significant differences between the responses collected in different semesters unless otherwise noted. The

[2] https://slack.com/

majority of the respondents (94%) were freshman with 5 outlying juniors. 92% of the respondents were Computer Science majors; the remaining students majoring in Information Science or Digital Narrative and Interactive Design. 76% of the responses were male, and 24% were female. The self-efficacy ratings were high, with an average score of 3.87 (range from 1 to 5, SD = 0.84). As far as the messaging tool usage, approximately half (39 of 83) of respondents used Slack. 82% of users expressed a certain level of familiarity with chatbots, and almost all the students (96%) expressed interest in a chatbot that can help them with course-related activities.

To understand students' current study behaviors, we asked a series of questions about their challenges and the use of various academic resources available to them. 38% of students in the Fall 2019 semester agreed to having difficulty managing time for a given assignment, while the number became significantly higher (72%) for students who were surveyed in Spring 2020 ($t(81) = 3.267$, $p<0.05$). A large proportion (85%) of freshmen tend to get help from school resources when they are in trouble. In terms of resource usage, students were more familiar with class-based resources such as *instructor office hours, TA office hours and Stack Overflow (a CS online Q&A site)*, with students surveyed in Spring 2020 semester expressing significant higher familiarity ($p<0.05$ for all three resources above). However, the majority of the students were unfamiliar with the department tutoring center, with only 18% expressing at least moderate familiarity. Similarly, familiarity was also low for online resources like MOOCs and educational software (52%), as well as classmates who had previously taken the course (45%). These results suggest first-year students have not yet fully recognized and utilized the vast academic and online sources that could help their studies, especially in their first semester.

As a part of the students' perceptions of a chatbot intervention, we asked students to rate the usability of several potential features of the chatbot as shown in Table 1. All the features were rated fairly high, with an average score above 3.8 out of 5. The students' higher rated feature tended to involve external support, such as offering tips from more experienced students, connecting users with a tutor, and recommending academic resources. Other features that focused on improving behavior or study skills, such as goal and time management or behavioral feedback, were perceived as useful but slightly less than the former. In an open-ended question where students were invited to brainstorm chatbots' potential functionalities, some students favored "online chat with the TA", suggesting the potential of using a chatbot as a communication channel between students and tutors. Overall, our target users responded positively regarding our proposed chatbot features.

Table 1. Proposed Design features and perceived usefulness score (ranging 1 to 5) from the in-class freshman survey (n = 83)

ID	Potential feature of the chatbot	M	SD
1	Computer Science tips from experienced and graduated students	4.28	0.80
2	Connecting me to a tutor or advisor when I really need help	4.24	0.78
3	Recommendations for academic resources available to me as a student	4.06	0.78
4	Reminders for completing my academic tasks on time	4.00	0.99
5	Help in breaking up large tasks into smaller, more manageable goals	3.96	0.82
6	Feedback for how well I am managing my time and practicing good study habits	3.95	0.97
7	Assistance in managing my times for assignments and academic tasks	3.86	0.93

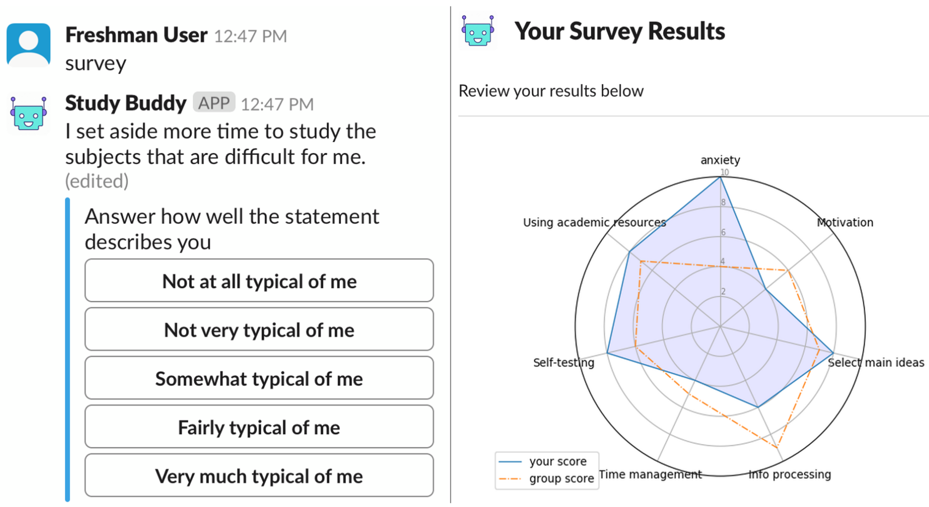

Fig. 1. The survey interface (left) and simulated results displayed using a radar chart (right)

4 STUDYBUDDY: A CHATBOT FOR EFFECTIVE STUDY HABITS AND SKILLS

The preliminary interview with the CS tutors revealed critical challenges students are facing, which brought to mind potential interventions for the chatbot. Following the interview, the survey further investigated these challenges and confirmed that our target users' proposed chatbot features were acceptable. To further test the feasibility and usability of these features, we designed and implemented a prototype that contains six features with the highest rated usefulness (Table 1). Among the six features, **#1 insider tips, #4 reminders, #5 task breaking-down** and **#6 study habit feedback** were implemented as a high-fidelity prototype. The remaining two features, **#2 connecting to the tutor** and **#3 recommending academic resources** were demonstrated as low-fidelity storyboards (as shown in Table 3). We then used these prototypes to conduct our usability evaluation and provide design recommendations related to behavioral intervention. The core technology of our chatbot, *StudyBuddy*, was built using DialogFlow [3] and later integrated into the Slack API [4]. We describe the features of the chatbot in the following subsection:

4.1 Description of Features

4.1.1 ***Insider tips:*** This feature periodically sends tips which we collected from experienced and graduated CS majors. This feature is grounded in the theory of situated learning which argues that succeeding in the context of an academic environment is unique and those who have been part of such a community could provide valuable insight and mentorship [2]. We designed two types of tips: *functional tips*, which are computer science task-related suggestions, e.g., related to debugging and writing pseudo code, and *motivational tips* which provide psychological encouragement to the students, e.g., encouraging a user to take a break or maintain good sleep habits. For the purpose of our usability evaluation, a student could type "tips" on the *StudyBuddy* chat interface and receive a tip as a response from the bot. A full list of tips are presented in Table 2.

[3]https://dialogflow.com/
[4]https://api.slack.com/

4.1.2 ***Study habit feedback:*** This feature aims to provide assessment and feedback of the students' study habits via surveys embedded in the interface. To evaluate the students' study habits, we adapted the questionnaire from the Learning and Study Strategies Inventory (LASSI) as it is a thoroughly researched and tested inventory [61]. The LASSI originally is a 10 scale 60 item inventory. Several studies mentioned the conceptual overlap and redundancy of the LASSI subscales [6]. So, we merged some scales e.g., motivation and attitudes and revised the questionnaire to be a 7 scale 14 item inventory as not to overburden a user in one interaction. Items could also be alternated in subsequent interactions to cover the entire survey. The 7 scales include: anxiety, information processing, selecting main ideas, time management, motivation, self testing and use of academic resources - all related to students' study habits and skills. The participants were presented with a series of 14 Likert style questions in the Slack interface to complete the survey. The survey is expected to take approximately two minutes to complete. The result of this survey allows students to track their study habits and identify their own limitations. This may inspire the students to make an effort to improve their shortcomings. This feature also provides valuable data from our users that could have further implications for user modeling, personalization, and appropriate intervention.

To trigger the survey, a student types "survey" in the *StudyBuddy* chat interface, as shown in Figure 1. The results are displayed to a student immediately after the survey is taken, along with the simulated averages of the students' peers. In an effort to break down the seven scales, a radar chart shows student's performance in each category. This feature, *study habit feedback*, was implemented utilizing Slack's Block Kit and API.

4.1.3 ***Task breakdown:*** Inspired by the peer-tutor interview described in section 3.1, this feature is designed to help students break down academic assignments into smaller segments. This allows the student to work on smaller tasks which are divided within the assignment duration. We designed a generic breakdown of a programming assignment: reading the description and understanding the problem, writing the pseudo code, writing the actual code, testing the code, and checking grading criteria of the assignment. This feature is implemented using DialogFlow (with nested intents and fall-back responses). To trigger this feature, a student could type "start with assignment" keyword on the *StudyBuddy* chat interface to receive guidance. An example of the dialogue is shown in Figure 2.

4.1.4 ***Scheduling and reminders:*** This feature addresses students' procrastination by encouraging students to think actively about time management. Our design anticipates the students will develop critical time management skills with continued use. We prototyped (Figure 2) a scheduler system that asks the students to input the deadline of their projects, dividing it into smaller milestones with corresponding reminder dates. Finally, the chatbot was expected to send reminders for completion of each sub-task of the assignment on the pre-scheduled dates. Like the *study habit feedback* feature, the interface was implemented utilizing Slack's API and Block Kit builder.

4.1.5 ***Low Fidelity Prototypes: Storyboard of the remaining two features***. The remaining two features involve more complex interaction and technologies (#2 connecting to the tutor and #3 recommending academic resources), thus we prototyped it using a storyboard method to proceed with the investigations related to our research questions. For a holistic narrative, the storyboards included the previous *Task breaking-down* and *Scheduling and reminders* features. To investigate how features and interaction complexity affect a user's perception, we designed three storyboards (e.g., Figure 3) with each focusing on different design choices and interaction complexity between a student and the bot. The characteristics of each storyboard are listed in Table 3 and fully described in the *Appendix*. The first storyboard, called *assignment starter*, contains a simple task breaking-down

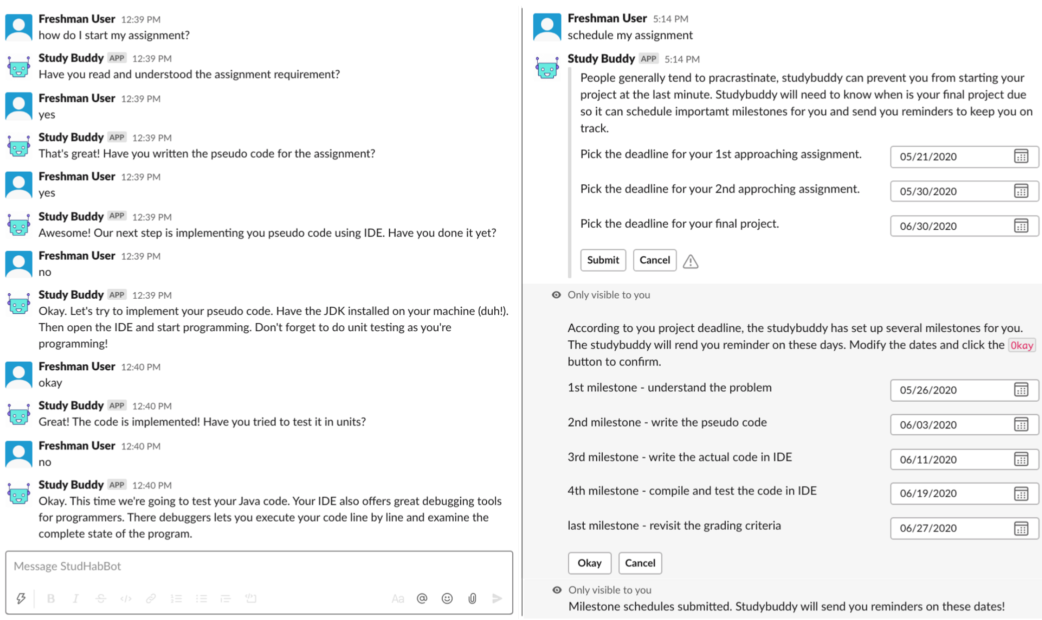

Fig. 2. StudyBuddy chatbot interface. Left: Partial dialogue of the task breaking-down feature, illustrated in storyboard 1. Right: the interface of the scheduling and reminders feature, illustrated in storyboard 2

feature. The second storyboard, the *assignment rhythm*, includes task breaking-down, schedule and reminders, and connecting to a tutor features. This storyboard emphasizes keeping students on track and encouraging active thinking in completing their assignment on time. The third storyboard, the *concept-learning facilitator*, was designed to proactively encourage students to ensure understanding of the concepts behind the assignment before starting their work. In addition to the task breaking-down, and schedule and reminders features, this storyboard also provides academic resource recommendations to improve their domain knowledge. We estimated the interaction complexity for these storyboards to be easy, moderate, and complex, correspondingly.

5 EVALUATION

After designing the aforementioned features and implementing prototypes and storyboard 1 & 2, our next step was evaluating its usability after integrating the features into the *StudyBuddy* chatbot. We first piloted the design through usability interviews with 8 users from two student groups, first-year and senior students. Using the feedback gathered from these interviews, we designed a usability survey to understand their expectations of usage on a larger scale (117 responses). We also conducted interviews with 5 course instructors to better understand educator perspectives on *StudyBuddy*. Gaining insights from multiple stakeholders proved to provide valuable feedback on how to maximize the benefits to our target users and how to promote a better collaboration among students, tutors, and instructors.

5.1 Usability Evaluation Through Student Interviews

5.1.1 ***Study procedure***. The student usability interview was conducted over two semesters. During the Fall 2019 semester, we recruited 4 computer science students (2 senior undergraduates,

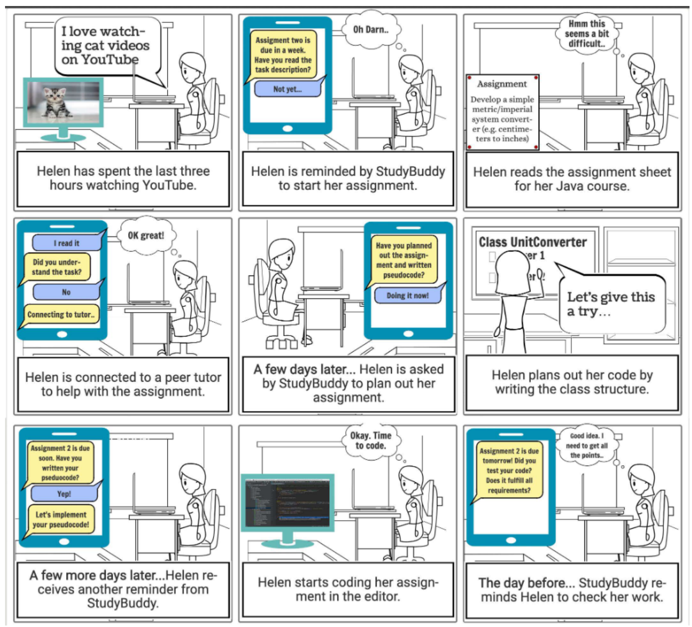

Fig. 3. An example of storyboard (#2) demonstrating the task breaking-down, reminders and connecting to the tutor feature

2 Ph.D. students; 2 male and 2 female) through direct contact and paid $10 gift cards for an average of one hour of their time. The usability interview was conducted in the lab where one researcher among us conducted the interview, while another was present to take notes. In the Spring 2020 semester, 4 Computer Science freshmen (1 male, 3 female) were recruited through campus email lists and social media. Each participant received $10 gift card as compensation. To increase the incentive, participants were informed they will be randomly selected to receive $100 cash after completing the study. The freshman study was moved online because of social distancing requirements from the COVID-19 pandemic. Our aim for this interview was to obtain qualitative feedback about the utility of *StudyBuddy*. The results from this study informed the future questions for the broader population.

In the Fall 2019 interviews, we tested the *insider tips*, *study habit feedback* implemented in *StudyBuddy* as well as three storyboards demonstrating the rest of our features. To interact with *StudyBuddy*, users were invited to join the Slack workplace where the chatbot was deployed and users were able to directly message *StudyBuddy* like with a regular user. A screenshot of a dialogue between user and *StudyBuddy* can be seen in Figure 2. In Spring 2020 interviews, we tested all features from the previous semester in addition to the *task breaking-down* and *scheduling and reminders* implemented in *StudyBuddy*. We used the same user study protocol as before except 1) adjusting the guidance to a virtual study setting 2) testing additional prototyped features included in storyboard 1 & 2.

During the usability study, there were three interaction tasks to be completed by students. The first task tested the usability of the *study habit feedback* feature, including quality and comprehensibility of the survey questions, interaction flow, and the visualization of the survey result; the second task involved testing the usability of *insider tips*, where we asked the students to rank all the tips according to their perceived helpfulness and explain their reasoning; and for the third task, we walked them through the different cells of the storyboard which showed the interactions between Helen, a hypothetical student character who has a programming assignment due soon. The participants had to understand the problem from Helen's perspective and see how the chatbot could benefit her. We asked a few semi-structured questions, e.g., *Do you think Helen will be helped by the chatbot? Why?*, *Are the interactions with the chatbot tedious? Why?*, *Among three storyboards, which one do you like the most? Why?*. We expected to identify elements that facilitate or hinder the intended behavioral change and factors that might influence their perceived usefulness or engagement with the chatbot.

We reference Fall 2019 senior participants as s1, s2, s3, s4 and Spring 2020 freshman participants as f1, f2, f3, f4 respectively in our following analysis. Table 2 and 3 shows the perceived usability of each tip and storyboard by all 8 participants. We summarize major findings from this study below:

5.1.2 ***Preference on domain-specific help and limiting interaction costs.*** Throughout the usability study, there were some broader takeaways and consensus about providing domain-specific help and limiting interaction costs. Our tips were clustered into two categories: motivational and functional, with the former attempting to inspire future studying and build rapport and the latter to provide more pragmatic help in domain-specific tasks. Table 2 indicates that among all tips we provided to freshman and senior students, functional tips, especially those that provided domain-specific information, were generally perceived as more useful than motivational tips (*"they are honestly looking for tips and tricks to think simpler" (s1)*). Likewise, in the breaking down tasks features, students appreciated domain-related processes like pseudocoding and some desired additional information on how to perform these subtasks. Interaction costs, the total effort a user must deploy with the chatbot to reach a goal, were also a concern with many features. The system could be tedious if it asks too many questions before being able to interact with the participant with relevant information. As far as the study habit feedback, participants overall found the feature useful and were able to interpret the results viewed in a radar chart, but desired a longer interval while conducting this self-evaluation in relation to the semester (beginning, middle, and end).

Table 2. All functional and motivational tips and their usability ranked by each participant; the ranking order is indicated by color shades (the darker the color, the higher position the tip was ranked. Order 1-3 was colored as dark green, 4-7 as mild green and 8-11 as light green)

Overall ranking	Category	Tip content	Ranking by each participant (The darker indicates higher rank)							
			f1	f2	f3	f4	s1	s2	s3	s4
1	Functional	If you're stuck with something, try visualizing, pen and paper!	1	1	6	3	2	1	3	1
2		Before you write the program, try to visualize the entire idea in your mind, come up with the main cases, write the algorithm, have a pseudo code. Then, your programming will be faster and have less bugs.	2	3	1	2	3	2	6	2
3		Always write code in incrementally functional bits.	5	2	3	4	9	6	4	6
5		It is always a good idea to write functions in your program	6	7	5	5	6	4	5	5
6		To debug a program you can place print statements (a quick and dirty way).	8	6	7	1	7	3	8	4
9		The debugger is your friend.	7	5	8	7	8	9	7	3
10		Algorithms is what makes us separate from others, try to be good at them!	9	8	9	6	10	8	9	7
4	Motivational	If you're studying late in the night, make sure to get some sleep before the test.	4	4	10	8	1	5	1	9
7		....it's okay to take a break.	10	10	2	9	4	7	2	8
8		When the going gets tough, the tough gets going!	3	9	4	10	5	10	10	2
11		Coffee is your second friend.	11	11	11	11	11	11	11	10

Table 3. Theme, features and interaction complexity of three storyboards and the usability ranked by each participant, with darker shading indicating higher perceived usefulness.

Storyboard #	Theme	Features included	Complexity	Ranking by each participant							
				f1	f2	f3	f4	s1	s2	s3	s4
1	Assignment starter	Breaking-down tasks	Easy	2	3	3	3	3	1	2	1
2	Assignment rhythm	Breaking-down tasks, reminders, connecting to tutors	Moderate	1	2	1	1	2	2	1	3
3	Concept-learning facilitator	Breaking-down tasks, recommending resources	Complex	3	1	2	2	1	3	3	2

Users also expressed similar concerns about too many interactions from the chatbot (e.g., the 3rd storyboard), favoring simplicity that didn't burden a user.

5.1.3 ***Variations among individuals.*** Despite the general consensus across chatbot features and users, we also observed a variation in user preference based on their personality, proficiency, and existing habits, indicated by differing perceived usability for tips and storyboards in Table 2 and 3. For example, some students found motivational tips unnecessary, remarking *"I already take a break" (f4)* as a response to tip #7 (as shown in Table 2), or not applicable to their problems: *"In CS, you can only achieve motivation if you succeed in a project" (s2).* Others seemed to value potential psychological benefits, noting that *"sometimes people are not motivated and feel guilty... taking a break might gain insight into the problem" (s1)*, or they valued potential rapport building: *"I will feel closer to the bot if it has a fun personality..." (s4).* Differences in proficiency also were apparent, in the feature to help break down tasks, one student expressed a lack of knowledge in how to write pseudocode (f4), while others believed the feature might *"hold their hand too much" (s2)* or become tedious with increased proficiency (f1,f4). Participants also showed preferences in social comparisons, for while some acknowledged the usefulness of comparing to classmates others preferred internal comparisons with their past results. In terms of existing habits, one participant pointed out their terrible time management habits, expressing a desire of being kept on track by our technology (s1). Other participants who already cultivated satisfying habits perceived these features to be less applicable.

5.1.4 ***Importance of contextual knowledge.*** Participants also acknowledged the importance of context for interacting with the chatbot. Tips were perceived more relevant if the topic was aligned with the student's current tasks and more helpful if received before an upcoming exam or assignment deadline. Likewise, a participant noted CS coursework varies with each semester and may only be a small portion of the overall work (*"I only had [CS courses] two days a week" (s2)*). For breaking down tasks, participants agreed that interactions should be specific enough to help them start the task and favored the storyboard that had existing contextual knowledge about their assignment (f2, s1). This issue of context also pertained to concerns for establishing common ground with the chatbot, one participant noting *"it's difficult to explain to the chatbot what we are doing in the class" (f4).* Overall, context was observed to be a variable that would impact the usefulness of several proposed features.

5.2 Usability Survey Understanding Expectations

Building on the above evaluation, we sought to understand stakeholder expectations towards our proposed design at a larger scale. To accomplish this, we produced a video [5] demoing our high fidelity prototype, incorporating wizard-of-oz elements, including all six features described above, for a holistic demonstration of the technology. Surveys were distributed in four computer science courses (two introductory, two upper level) at our university during the first few weeks of Fall 2020 semester. The usability survey, included in the *Appendix*, consisted of a 3-minute video demonstration of the chatbot, a technology acceptance model (described below), demographic information, and specific feature related questions. As an incentive, students were given the chance to enter their contact information to be randomly selected for a $25 reward.

5.2.1 ***Participants and study procedure***. A total of 117 valid responses were collected in Fall 2020 from all undergraduate levels with 63% freshmen, 9% sophomore, 11% junior, 16% senior, and 1% other. The average completion time was 12.2 minutes (median was 9.8 minutes). At the time of survey distribution, freshmen had just experienced their first few weeks of college and, unlike non-freshmen, had yet to establish a routine for college life. Because of this sharp difference, our comparison collapsed all non-freshman students into one group, which consists of 37% of the sample. It should be noted that this university was supporting a hybrid model of learning as a result of COVID-19 that allowed students to attend courses in person or online, permitting switching at any time. Students were allowed back on campus and dormitories were open.

The survey participants were comprised of 73% Computer Science, 10% Information Science, and 17% in other STEM fields (such as Physics or Computational Biology). In terms of gender, 70% of responses are male, 28% are female, and 2 participants identified themselves as non-binary. In order to understand to what extent first-year students might struggle to adapt to the pace of college curriculum, we asked a series of questions investigating existing experience with computer science learning. We examined these responses for the 74 first-year students. The majority had high school class (86%) and AP exam or class (70%) experience, 43% selected self-taught, 27% had taken online courses, 18% had experience from other college-level sources, and 5% had professional experience.

5.2.2 ***Technology acceptance of StudyBuddy***. Our technology acceptance survey was informed by two extensions to the Unified Theory of Acceptance and Use of Technology (UTAUT) that targeted mobile learning technologies [8, 60]. The survey consisted of 6 constructs: self management of learning (SML), trust, effort expectancy (EE), performance expectancy (PE), satisfaction, and behavioral intention (BI). Each construct consisted of 3-4 items for a total of 21. We used 5-point likert scale to measure each response item. To verify our constructs we performed Confirmatory Factor Analysis (CFA). The comparative fit index (CFI) = .928, the Tucker-Lewis fit index (TLI) = 0.914, and the RMSEA = .074 indicated a good model fit.

We then conducted multiple linear regression ($F(5, 111) = 31.05$, $p < .001$, $R^2 = .583$) on the averages for each construct with behavioral intention as the response variable (coefficients shown in table 5). As anticipated by the UTAUT model, there was a significant effect with performance expectancy and satisfaction, however effort expectancy and trust were not significant predictors. Most notably, there was a significant negative effect for self management of learning where one point increase was associated with a .299 decrease in behavioral intention.

We also investigated differences in gender and current class standing, breaking the latter into first-year and upper level students. Due to concerns regarding the normality of the data after conducting a D'Agostino-Pearson test, we chose to conduct a nonparametric test. For class standing, a Mann-Whitney U test revealed no significant differences, however a notable trend was present in

[5] System demo video: https://youtu.be/bLlDL5UCMeI

Table 4. Multiple Linear Regression Independent Variable Constructs

	SML	EE	PE	Satis.	Trust
β	-0.299	-0.229	0.460	0.58	0.237
t	-3.16	-1.37	3.39	.4.35	1.66
p	.002	.174	.001	.000	.099

behavioral intention of first-year (Mdn = 3.67) and upper level (Mdn = 3.0) students (U = 10.5, n1 = 73 n2 = 44, p = 0.055 two-tailed). For male (n=82) and female (n=33), a two-tailed Mann Whitney U test revealed significant differences for all constructs except self management of learning.

Table 5. Comparison of UTAUT constructs between Male and Female

	SML	EE	PE	Satis.	Trust	BI
U	-1548	959.5	941.5	919	732.5	889
Mdn	(4.00, 3.75)	(4.00, 4.00)	(4.00, 4.00)	(3.67, 4.00)	(3.67, 4.33)	(3.00, 4.00)
IQR	(1.19, 1.25)	(0.75, 0.50)	(0.75, 0.50)	(1.00, 1.33)	(1.00, 0.67)	(2.00, 1.33)
p	.226	.013	.010	.007	.000	.004

Looking specifically at behavioral intention, which were averaged from 3 similarly phrased items (scale from 1 to 5), we decided to break users into three groups of low (x <=2.5), neutral (2.5 < x < 3.5), and high (x >= 3.5) in an effort to better understand how factors influenced each group. Roughly half (49.6%) of all students expressed high BI with some differences in gender and class standing. While two thirds of all females expressed high BI, only 41.5% of all males reported the same. Similarly, 56.2% of all freshmen fell into the high BI group as opposed to 38.6% of upper level students. Overall, most students fell into the high (n=58) group, and the remaining respondents were neutral (n=26) and low (n=33) in BI. Contrary to the regression analysis, a Kruskal–Wallis test between groups did not differ significantly in self management of learning, although a Mann–Whitney U test revealed students with low BI (M=3.95) did differ significantly with students with high BI (M=3.63) in SML (U = 1153.5 p < .001). Likewise, trust was shown to significantly differ between groups of students (H(2) =34.62, p < .001) unlike the regression analysis. Further analysis revealed the neutral BI group (M=3.69) expressed significantly lower trust than the high BI group (M=4.29) (U = 1153.5 p < .001). Overall these findings suggest that self management of learning was a factor for expressing low behavior intention, but not so much for neutral behavior intention. Instead, these students on the fence had significantly different levels of trust than those with high BI.

5.2.3 ***Student perceptions of chatbot features.*** To understand which aspects of our prototype students valued most, we asked participants to rank the six features based on perceived helpfulness (results shown in Figure 4). Students were also asked to explain their reasoning for their highest and lowest rated features. Students seemed to prioritize planning and organization features, as breaking down tasks and the scheduler/reminder were ranked highest. This is understandable as students may feel overwhelmed and lose track of their numerous assignments, projects, and other responsibilities from various classes. A female first year student remarked, *"With lots of things on my plate it's hard to remember what I need to do."* Additionally, these features are especially crucial under a global pandemic as students adapt to a radically different learning environment. A male sophomore commented, *"With virtual learning, it's easy to get lost and not notice assignments posted*

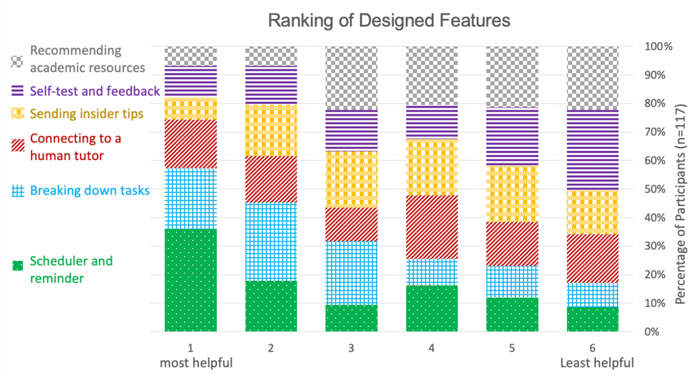

Fig. 4. Features ranking from usability survey responses, features dominant on the bottom left was the most helpful, on the top right was the least helpful.

on one of the many sites used." Overall, features that help organize time and balance the amount of work were prioritized by most students.

The next two highest ranked features involved a collaborative element: connecting to a tutor and receiving insider tips from experienced students. Students valued communicating with members of their academic community, whether directly with a tutor or indirectly from those who already went through similar experiences. These preferences appeared motivated by concerns about purely automated assistance. Two sophomore males expressed concerns over a general lack of ability in the bot: *"...discussing with an [AI] when I'm really confused would make me not want to use it anymore"* and *"...(human tutor) is an actual person that can understand nuanced human problems better."* A male freshmen more specifically referenced the domain of learning, stating that *"questions regarding programming issues are very specific and would be difficult for a computer bot to answer in a helpful way."* Automated connections were also seen as potentially impractical, but not as a technology deficiency. For instance, on the tips feature, a female freshmen remarked *"Graduated students are the best resource for navigating college, but are hard to get a hold of..."*

The second lowest rated feature was self-assessment and feedback of study habits. Among those who ranked this feature at the bottom, 15 out of 29 students expressed they already felt good about their own study habits and did not desire feedback nor have the intention to change their existing habits. Some students expressed an indifference or lack of trust in the feedback. For instance, a male freshman stated, *"I don't care what a bot thinks about my life."* Apprehension to engaging in the assessment feature could also be related to broader concerns about sharing personal data. Specifically, half of students (59/117) were not willing to share their results to external stakeholders. For students who were willing to share, they prefered to share their results with faculty and staff such as the school or department (62%) or academic advisors (45%), rather than other relations such as friends, peer mentors, or family. Only one other feature was ranked consistency lower, and this was recommending academic resources. Most students believe existing tools (such as Google) or other sources like instructors, are enough to help them obtain resources and recommendations. For instance a male sophomore stated, *"the professors likely provided enough resources to begin with, if a deadline is approaching, the last thing a student wants to do is read something that likely won't be exactly what they need."* While these two features were ranked lower by users, they still could have potential value. Collecting data of study habits could provide crucial insights into universities and academic advisors. Likewise, some form of recommendation could still have merit, and given the preference for humans as a source of information, perhaps *StudyBuddy* could facilitate recommendations from peers.

Last, we looked at differences in rankings between males and females and between new and experienced students. A two tailed Mann–Whitney U test revealed that the ranking for *connecting students to human tutors* was significantly different between the new (Mdn = 4) and experienced (Mdn = 3) student populations (U = 2010, P < 0.05). Freshmen ranked this feature significantly lower, for while 55% of the upper class ranked this feature in their top 3, only 40% freshman ranked the same. Likewise, the breaking down tasks feature was ranked significantly higher by females (Mdn = 2) than males (Mdn = 3) (U = 1712, P < 0.05). Perhaps more notably, 33% of females ranked this feature as No. 1 instead of only 16% males. While only one feature differed significantly for each comparison, these findings do suggest that gender and experience play a role in the appraisal of features.

5.3 Usability Evaluation through Instructor Interviews

5.3.1 ***Study procedure***. To evaluate *StudyBuddy* from a course instructor's perspective, we conducted semi-structured interviews with five instructors who teach courses at the CS department. The interviews of roughly 30 minutes were conducted remotely to practice social distancing due to COVID-19. Two of the five instructors were teaching fellows, advanced Ph.D. students who serve as course instructors, another a teaching assistant professor, and the remaining two were associate professors. All of participants had served as instructors or TAs for at least five years, covering a wide range of CS courses from introductory to higher level courses. Courses taught typically had 30 to 40 students, with some enrollments as high as 100. We invited the instructors to watch the previously described video demo of *StudyBuddy*, highlighting its features before the interview. During the interview, we asked questions regarding the instructor's teaching experience, class management tools, and their opinion on the design features of *StudyBuddy*. We refer to these five instructors as I1, I2, I3, I4, I5.

5.3.2 ***Contrasting perceptions between instructors and students***. Part of our interview involved a feature ranking task, in which we observed faculty rankings often differed from students. For example, while *connecting to human tutors* and *recommending academic resources* were not the most valued by students, instructors ranked them as the most helpful features. The instructors thought *connecting to tutors*, a feature that involves the most human interaction, was the most flexible way to help students with their coursework and address their questions. Instructors often situated features in the context of their courses, noting that *recommending academic resources* could be improved by displaying contents that were made during the lecture, instead of just results from a search engine. Similarly, instructors expressed unique concerns in the context of their role as educators. They felt an overly specific breakdown of tasks could lead to a dependency on the technology and would prevent development of long-term study habits. Instead, I5 desired scaffolding the task, stating the chatbot should *"ask students to manage projects themselves, which is an important aspect to prepare for industry."* Additionally, instructors brought up the possibility of cheating, leading them to favor less the *sending insider tips* feature as they thought this would provide a mechanism for abuse. For example, I1 said, *"It could be used in a negative way. Teachers don't change course content, it's easy to cheat."* The instructors suggested alternative ideas related to the tips that might be beneficial for the students, For example, peer-sourced and discussion-sourced tips from discussions among students on forums like Piazza [56] could prove to be useful for other students who missed out or did not participate in those discussion. Additionally, the tip feature could focus more on motivational phrases as noted by I1: *"(For) students who are doing good but with no acknowledgement, (tips can serve as) a positive reinforcement, (it's) nice to hear it from an* *external entity"*. Overall, teaching experience and perspective as an instructor led to pedagogical and course management concerns which impacted their rankings.

5.3.3 ***Will chatbot reduce teaching burden?*** Other questions focused on *StudyBuddy*'s ability to help their teaching and lessen their burden. Instructors responded that instead of entirely replacing the learning experience of an in-person teaching setup, *StudyBuddy* can be used to aid teaching by recommending appropriate resources and facilitating communication and assistance. For example, *StudyBuddy* might respond to logistical questions regarding the course on behalf of instructors and TAs. In this manner, instructors and TAs will have more time for important questions regarding the class material that students may have difficulty with. According to the instructors, systems like *StudyBuddy* should be built with autonomy and packaging, to reduce the learning curve associated with new systems and help the instructors to easily incorporate them in their course. Overall, instructors seemed most comfortable with the chatbot handling simpler tasks, like basic course information or mediating communication, but still felt *StudyBuddy* had the potential to integrate into the courses and alleviate some of their burdens.

6 DISCUSSION

Behavioral change takes time and typically requires the adoption and use of persuasive technology for a prolonged period. We identified barriers that might hinder a student's use of a persuasive chatbot like *StudyBuddy*. These barriers inform our design implications and recommendations related to chatbot-based interventions for behavior. We hope our findings will inspire a continuing discussion within the CSCW communities as to the future trajectory of conversational education technologies.

6.1 Trust and Privacy

Trust in the system was found to be difficult to obtain, which could deter users or result in abandonment. As indicated in our usability survey, student's level of trust differed significantly among students with high, neutral, and low behavioral intention. Much of students' concerns with trust seemed related to the chatbot's ability to understand user's input and perform the designated task. This aligns with prior findings that many intelligent agents struggle to convey genuine intelligence and keep users engaged over repeated encounters [27, 41]. Part of this is due to the technical limitations that do not fulfill students' expectations toward the agents. Previous attempts at addressing this problem applied personalization [27] or added diversity to the interaction [12]. Similarly in our context, to keep the student engaged and build trust with the agent, the chatbot needs to incorporate advanced NLP techniques that empower question answering [16] or chitchat conversations [59]. While demonstrating conversational intelligence can help build trust, both overly complex and excessive interactions with the chatbot might disengage the student. Among the three storyboards, the majority of the students preferred the second scenario with moderate interaction complexity. A simpler interaction of a chatbot might not stimulate the thought process of a student while adding too much complexity might frustrate students and increase interaction costs. There needs to be a design threshold of questions being asked to students before providing constructive suggestions which we address further below.

Beyond trust in a chatbot's ability, there were also issues with trusting a dialogue agent with personal data. While instructors found it helpful to gather student behavior data to inform their teaching, many students prefer limiting data sharing due to privacy concerns. All five instructors we interviewed wanted to learn what topics and concepts their students are struggling with, which could be summarized semantically from the dialogues students had with the chatbot. Yet despite the potential benefits, only one third of the students (39/117) were willing to share their study habits survey data with the school/department. As a senior male student noted, *"I don't mind if the data is kept and shared internally within the system, but I don't like when other people can see my habit specifically out of everyone else."* As noted by one of the instructors (I5), collecting behavioral data itself might cause undesirable anxiety or compromise students' honesty when interacting with the bot. Such challenges further echo issues in privacy research of how to carefully leverage the valuable personal data collected to support users while not violating personal privacy.

6.2 Immediate Help vs. Long-term Sustainable Support

Our interviews with faculty highlighted the importance of considering immediate vs long-term support. Instructor's valued the ability of features to provide sustained support that helped students build lasting skills. Alternatively, students were more concerned about the present, as one senior stated *"students are simply looking for tips and tricks that make their life easier."* However, while favored features like scheduling and task management can be immediately useful, as students form their own study habits and problem-solving skills the use of these features would decline. This is not a new problem. Smart technologies, wearables, and persuasive technologies that seek integration with daily life are prone to high abandonment rates. In the health wearable setting, one study revealed that one third of users in the US who purchased wearable self-tracking products stopped using it after only six months of use [40]. Studies have also sought to identify key issues that prevent these technologies from sustained use [10, 39]. Identifying how to inspire continuous engagement of our chatbot is a critical aspect of habits formation which requires a longer period of time. The key element of the chatbot here is to help new students by articulating what effective study behaviors are (such as breaking down a large project) and facilitating practice over time. In designing for long-term support, these tools must adapt to changing needs as a student's ability and experience grow. Additionally, the inclusion of features that will not degrade over time, for instance those that integrate into existing courses and the academic community, will provide lasting value for users.

6.3 Gender and Individual Differences

Our findings from our evaluation suggested that gender had a significant impact on the perception of *StudyBuddy*. Most notably, females expressed higher ratings across the board for technology acceptance, including higher trust and behavior intention. This leads to a broader question of why a study aid would appeal more to females, and if this might be attributed to their current state of being an underrepresented group in computer science education. Past research has noted females to express lower computer self-efficacy [3, 4], perhaps making a study aid more enticing. As evident in our sample, females are also severely outnumbered in computer science, leading to potential feelings of isolation. Pooja Sankar reported isolation as a female computer science student as her motivation for creating the popular discussion tool Piazza [56]. Piazza's exhaustive study of over 1 million students found females ask more questions than males and are more likely to make use of anonymity tool to ask questions. *StudyBuddy* could offer a less intimidating medium to ask questions, gather information, and communicate with instructors and TAs.

Other individuals differences, like self-management of learning, also had a significant impact. Students that already had a high appraisal of their study skills expressed low behavior intention. This is somewhat intuitive that a study aid would appeal less to those who already possess good study habits and skills; however, this also presents important design considerations. By specifically targeting a subset of students, such as those underrepresented or having lower self-management of learning, designers can prioritize making the technology inclusive to those perceived at higher risk of attrition. Alternatively, a universal design approach might try to include features for those more confident in their abilities. Both approaches have drawbacks as focusing on one target population could lead to excluding others. Further, a design that attempts to cater to all possible types of students may make the technology overall less effective.

6.4 Personalizing the Chatbot Experience

From both students' and course instructors' evaluation of *StudyBuddy*, a 'one-size-fits-all' approach to the design is unlikely. In our context, we observed that the perceptions of the usability of a chatbot supporting students' study habits differed by the students' gender, class standing, prior experience, and self-management of learning. Given this evidence, personalized chatbot may be the best approach to improve students' study habits. Prior research has emphasized the importance of personalization in pedagogical agents regarding learner's situational interest [47], empathy [41], and relationship between the agent and learners [62]. To personalize students' experience, for example, the chatbot could customize the tips suitable for a particular type of student, e.g., sending more motivational tips to students with low motivation. The instructor interviews also confirmed such design recommendations. For example, if the system can track students' state concerning the class assignment/project, it can tailor features to the specific students' need, e.g., providing scaffolding to the students' who are struggling to break down tasks vs no aid for students who have already demonstrated ability and persistence. Additionally, personalization can augment dialogue, as techniques like user modeling have already been applied to generate personalized responses in chatbots [31]. For example, if a student is more responsive to a chatbot's questions, then a question-based pattern of dialogue can be adopted. Similarly, if a student has difficulty adhering to a study schedule, the chatbot can send more frequent reminders with more encouraging prompts. Given these additional features, personalization could be a key factor for helping these chatbot systems be adopted for a period of time long enough to cause a behavior changes, and ultimately, form good study habits.

6.5 Design for a Context-Aware Chatbot

To provide relevant and timely support to the students, the context of interaction needs to be taken into account. Literature notes that shared mental models or common ground through conversations increase both social relationships and learning [32]. Although context is heavily researched in IoT and wearables when conducting intervention [37, 48], context-aware education technologies remain underdeveloped. Furthermore, our usability survey found that designing a context-aware chatbot was rated to be of the highest concern for students[6]. Our recommendation is to integrate data collected by a university or department with much of this information residing in a learning management system. These systems like Canvas[7] and Moodle[8] provide REST API's to access student assignments, activity and materials in a course, offering a means of integrating context into a chatbot. In addition to an LMS, instructors could provide a source of context, allowing them to fine-tune specific features according to their class needs and expectations.

This surrounding knowledge can contribute holistically to the chatbot, allowing for synergy between the different features offered by *StudyBuddy* to create a system that could understand the students' needs from multiple angles. For instance, the chatbot could learn of potential motivational barriers based on student activity and tailor the interventions accordingly, making it a "personal study coach." Rather than generic information, *StudyBuddy* could select domain-specific tips that are relevant to their current task, recommend more relevant resources, and demonstrate awareness of upcoming exams or assignments. Features such as breaking down project tasks could

[6]Survey responses to "The bot isn't aware of my learning context" rated at 4.15 out of 5
[7]https://www.instructure.com/canvas/
[8]https://moodle.org/

adapt specifically to the assignment at hand. Contextual information would also further facilitate personalization and convey that *StudyBuddy* was a knowledgeable member of the collaborative learning community. The use of such information would allow a chatbot to facilitate communication between students, instructors, TAs, and tutors, expanding the tool's role to be an active agent in the learning environment.

7 CONCLUSION

In this work, we studied the emergent practices of a chatbot for effective behavioral change to support college students' study habits. Our usability research highlighted the important role context, gender, personalization, and adaptation play in this design space. We identified several barriers that potentially hinder students' engagement and long-term learning, proposing several design considerations.

We deployed a large-scale usability survey to address the small number of subjects in the usability evaluation due to university closures in response to the COVID-19 pandemic. The single-session nature of the usability evaluation produced valuable user feedback and the usability survey helped to better understand users' perceptions. However, neither provide the same level of rigor as a deployed, long term study which we aspire for future work. We recognized this limits the conclusiveness of our findings. While we believe a longer-term study is needed, we also believe insights were gained that can greatly inform that work, as well as other work in the community pursuing similar goals.

A persuasive chatbot for study habits and skills could bring broader influence on the college learning environment. First, freshman-year challenges are recognized by many universities (e.g., the early academic alert systems [21, 30, 57, 58]), but these systems require extensive involvement by the current faculty. The chatbot could reduce faculty's load by proactively detecting individuals who may be experiencing academic difficulty and refer them to appropriate support systems that tailor to meet student's issues. Second, the chatbot can provide valuable information to help the department make decisions on curriculum formation, teaching resource allocation, and more. Third, the chatbot has an opportunity to bring a positive impact on the increasing demand for distance education. With a lot of academic activities moved online due to COVID-19 [11], a chatbot could be an important tool to support new routines of study, e.g., scheduling an appointment with online tutors. Overall, study habits and skills are important constructs that directly affect academic success. Persuasive education technologies, such as chatbots, have great potential to influence the study habits of college students in their early academic years.

ACKNOWLEDGMENTS

We thank Dr. Matthew Burton for helping us distribute surveys and recruit participants in the class. We also thank all the peer-tutors, undergraduate, graduate students and course instructors who participated in our user studies in the Department of Computer Science at University of Pittsburgh. Finally, we thank Joseph Wiggins for his thoughtful proofreading and all the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] RV Belfin, AJ Shobana, Megha Manilal, Ashly Ann Mathew, and Blessy Babu. 2019. A Graph Based Chatbot for Cancer Patients. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, Coimbature,India, 717–721.

[2] Mordechai Ben-Ari. 2004. Situated Learning in Computer Science Education. *Computer Science Education* 14, 2 (2004), 85–100.

[3] Sylvia Beyer. 2014. Why Are Women Underrepresented in Computer Science? Gender Differences in Stereotypes, Self-Efficacy, Values, and Interests and Predictors of Future CS Course-Taking and Grades. *Computer Science Education* 24, 2-3 (2014), 153–192. https://doi.org/10.1080/08993408.2014.963363

[4] Sylvia Beyer, Kristina Rynes, Julie Perrault, Kelly Hay, and Susan Haller. 2003. Gender Differences in Computer Science Students. In *Proceedings of the 34th SIGCSE technical symposium on Computer science education (SIGCSE '03)*. Association for Computing Machinery, New York, NY, USA, 49–53. https://doi.org/10.1145/611892.611930

[5] Davide Calvaresi, Jean-Paul Calbimonte, Fabien Dubosson, Amro Najjar, and Michael Schumacher. 2019. Social Network Chatbots for Smoking Cessation: Agent and Multi-Agent Frameworks. In *2019 IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE, Thessaloniki, Greece, 286–292.

[6] Sandra Beermann Chacko and Mary E Huba. 1991. Validation of the Learning and Study Strategies Inventory With a Sample of Students in Nursing. *NACADA Journal* 11, 2 (1991), 5–13.

[7] S Chamundeswari, V Sridevi, and Archana Kumari. 2014. Self-Concept, Study Habit and Academic Achievement of Students. *International Journal of Humanities Social Sciences and Education* 1, 10 (2014), 47–55.

[8] Cheng-Min Chao. 2019. Factors Determining the Behavioral Intention to Use Mobile Learning: An Application and Extension of the UTAUT Model. *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.01652 Publisher: Frontiers.

[9] Marcia Roe Clark. 2005. Negotiating the Freshman Year: Challenges and Strategies among First-Year College Students. *Journal of College Student Development* 46, 3 (2005), 296–316.

[10] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-Tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 647–658.

[11] Emerson Collective. 2020. Supporting Students During COVID-19: Resources for Remote Learning. Retrieved May 31, 2020 from https://www.emersoncollective.com/covid19-resources-for-remote-learning

[12] Alexandre Coninx, Paul Baxter, Elettra Oleari, Sara Bellini, Bert Bierman, Olivier Blanson Henkemans, Lola Cañamero, Piero Cosi, Valentin Enescu, Raquel Ros Espinoza, et al. 2016. Towards Long-Term Social child-robot Interaction: Using Multi-Activity Switching to Engage Young Users. *Journal of Human-Robot Interaction* 5, 1 (2016), 32–67.

[13] Sunny Consolvo, David W. McDonald, and James A. Landay. 2009. Theory-Driven Design Strategies for Technologies that Support Behavior Change in Everyday Life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, Boston, MA, USA, 405–414. https://doi.org/10.1145/1518701.1518766

[14] National Science and Technology Council. 2018. *Charting a Course for Success: America's Strategy for STEM Education.* Technical Report. Office of Science and Technology Policy Washington, DC.

[15] Marcus Credé and Nathan R. Kuncel. 2008. Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance. *Perspectives on Psychological Science* 3, 6 (Nov. 2008), 425–453. https://doi.org/10.1111/j.1745-6924.2008.00089.x

[16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* (2018).

[17] Mateusz Dolata and Gerhard Schwabe. 2018. Don't be Afraid! Persuasive Practices in the Wild. *Computer Supported Cooperative Work* 27, 3 (Dec. 2018), 427–462. https://doi.org/10.1007/s10606-018-9330-4

[18] Benedict Du Boulay, Katerina Avramides, Rosemary Luckin, Erika Martínez-Mirón, Genaro Rebolledo Méndez, and Amanda Carr. 2010. Towards Systems that Care: a Conceptual Framework based on Motivation, Metacognition and Affect. *International Journal of Artificial Intelligence in Education* 20, 3 (2010), 197–229.

[19] Teresa Garcia Duncan and Wilbert J McKeachie. 2005. The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist* 40, 2 (2005), 117–128.

[20] Andrew Emerson, Fernando J. Rodríguez, Bradford Mott, Andy Smith, Wookhee Min, Kristy Elizabeth Boyer, Cody Smith, Eric Wiebe, and James Lester. 2019. Predicting Early and Often: Predictive Student Modeling for Block-Based Programming Environments. In *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Vol. 39. ERIC, 48.

[21] Johna Faulconer, Jayne Geissler, Diane Majewski, and John Trifilo. 2014. Adoption of an Early-Alert System to Support University Student Success. *Delta Kappa Gamma Bulletin* 80, 2 (2014).

[22] Justin Filippou, Christopher Cheong, and France Cheong. 2016. Combining The Fogg Behavioural Model And Hook Model To Design Features In A Persuasive App To Improve Study Habits. *arXiv preprint arXiv:1606.03531* (2016).

[23] Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (2002), 2.

[24] Brian J Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*. 1–7.

[25] Ashok Goel, Brian Creeden, Mithun Kumble, Shanu Salunke, Abhinaya Shetty, and Bryan Wiltgen. 2015. Using Watson for Enhancing Human-Computer Co-Creativity. In *2015 AAAI Fall Symposium Series*. AAAI 2015 Fall Symposium, Arlington, Virginia, 22–29.

[26] Philip J Guo. 2013. Online Python Tutor: Embeddable Web-Based Program Visualization for CS Education. In *Proceeding of the 44th ACM technical symposium on Computer science education*. 579–584.
[27] Olivier A Blanson Henkemans, Bert PB Bierman, Joris Janssen, Mark A Neerincx, Rosemarijn Looije, Hanneke van der Bosch, and Jeanine AM van der Giessen. 2013. Using a Robot to Personalise Health Education for Children with Diabetes Type 1: A Pilot Study. *Patient education and counseling* 92, 2 (2013), 174–181.
[28] Carmen Holotescu. 2016. MOOCBuddy: a Chatbot for Personalized Learning with MOOCs. In *RoCHI*. RoCHI, Iasi,Romania, 91–94.
[29] Chin-Yuan Huang, Ming-Chin Yang, Chin-Yu Huang, Yu-Jui Chen, Meng-Lin Wu, and Kai-Wen Chen. 2018. A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion. In *2018 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, Bangkok, 1791–1795.
[30] Sandeep M. Jayaprakash, Erik W. Moody, Eitel J. M. Lauría, James R. Regan, and Joshua D. Baron. 2014. Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics* 1, 1 (May 2014). https://doi.org/10.18608/jla.2014.11.3
[31] Dipesh Kadariya, Revathy Venkataramanan, Hong Yung Yip, Maninder Kalra, Krishnaprasad Thirunarayanan, and Amit Sheth. 2019. kBot: Knowledge-enabled Personalized Chatbot for Asthma Self-Management. In *2019 IEEE International Conference on Smart Computing*. IEEE, Washington, DC, USA, 138–143.
[32] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human–Computer Interaction* 19, 1-2 (2004), 61–84.
[33] Saskia M. Kelders, Robin N. Kok, Hans C. Ossebaard, and Julia EWC Van Gemert-Pijnen. 2012. Persuasive System Design Does Matter: A Systematic Review of Adherence to Web-Based Interventions. *Journal of Medical Internet Research* 14, 6 (2012). https://doi.org/10.2196/jmir.2104
[34] Juho Kim, Philip J Guo, Daniel T Seaton, Piotr Mitros, Krzysztof Z Gajos, and Robert C Miller. 2014. Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos. In *Proceedings of the first ACM conference on Learning@Scale*. 31–40.
[35] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 881–894. https://doi.org/10.1145/3196709.3196784
[36] Ilya Kreynin, MohammedShabbar Manek, and Chirag Variawa. 2019. Creating a Virtual Chatbot to Scaffold Skills Development in First-Year Engineering Education. *Proceedings of the Canadian Engineering Education Association (CEEA)* (2019).
[37] Nicholas D Lane, Sourav Bhattacharya, Petko Georgiev, Claudio Forlivesi, and Fahim Kawsar. 2015. An Early Resource Characterization of Deep Learning on Wearables, Smartphones and Internet-of-Things Devices. In *Proceedings of the 2015 international workshop on internet of things towards applications*. 7–12.
[38] Randy A LaRose. 2009. The Relationship between Religiosity and Educational Pursuit and Perception. *All Graduate Theses and Dissertations* (2009).
[39] Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 635–646. https://doi.org/10.1145/2750858.2804288 event-place: Osaka, Japan.
[40] Dan Ledger and Daniel McCaffrey. 2014. Inside Wearables Part 1: How Behavior Change Unlocks Long-Term Engagement. (2014). https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3
[41] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic Robots for Long-Term Interaction. *International Journal of Social Robotics* 6, 3 (2014), 329–341.
[42] Kai Lukoff, Taoxi Li, Yuan Zhuang, and Brian Y Lim. 2018. TableChat: Mobile Food Journaling to Facilitate Family Support for Healthy Eating. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.
[43] Carol Maher, Jillian Ryan, Christina Ambrosi, and Sarah Edney. 2017. Users' Experiences of Wearable Activity Trackers: a Cross-Sectional Study. *BMC public health* 17, 1 (2017), 880.
[44] John Matthews, Khin Than Win, Harri Oinas-Kukkonen, and Mark Freeman. 2016. Persuasive Technology in Mobile Applications Promoting Physical Activity: a Systematic Review. *Journal of Medical Systems* 40, 3 (Jan. 2016), 72. https://doi.org/10.1007/s10916-015-0425-x
[45] Joel McFarland, Bill Hussar, Jijun Zhang, Xiaolei Wang, Ke Wang, Sarah Hein, Melissa Diliberti, Emily Forrest Cataldi, Farrah Bullock Mann, and Amy Barmer. 2019. The Condition of Education 2019. NCES 2019-144. *National Center for Education Statistics* (2019).
[46] Carolina Mega, Lucia Ronconi, and Rossana De Beni. 2014. What Makes a Good Student? How Emotions, Self-Regulated Learning, and Motivation Contribute to Academic Achievement. *Journal of educational psychology* 106, 1 (2014), 121.

[47] Joseph E Michaelis and Bilge Mutlu. 2019. Supporting Interest in Science Learning with a Social Robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, 71–82.
[48] Vivian Genaro Motti and Kelly Caine. 2016. Smart Wearables or Dumb Wearables? Understanding How Context Impacts the UX in Wrist Worn Interaction. In *Proceedings of the 34th ACM International Conference on the Design of Communication*. 1–10.
[49] Dina Fitria Murad, Adhi Gustian Iskandar, Erick Fernando, Tica Shinta Octavia, and Deryan Everestha Maured. 2019. Towards Smart LMS to Improve Learning Outcomes Students Using LenoBot with Natural Language Processing. In *2019 6th International Conference on Information Technology, Computer and Electrical Engineering*. IEEE, Semarang, Indonesia, 1–6.
[50] Seyed Mehdi Nasehi, Jonathan Sillito, Frank Maurer, and Chris Burns. 2012. What Makes a Good Code Example?: A Study of Programming Q&A in StackOverflow. In *2012 28th IEEE International Conference on Software Maintenance (ICSM)*. IEEE, 25–34.
[51] Kyo-Joong Oh, Dongkun Lee, Byungsoo Ko, and Ho-Jin Choi. 2017. A Chatbot for Psychiatric Counseling in Mental Healthcare Service based on Emotional Dialogue Analysis and Sentence Generation. In *2017 18th IEEE International Conference on Mobile Data Management*. IEEE, Daejeon, South Korea, 371–375.
[52] Rita Orji and Karyn Moffatt. 2018. Persuasive Technology for Health and Wellness: State-of-the-art and emerging trends. *Health Informatics Journal* 24, 1 (March 2018). https://doi.org/10.1177/1460458216650979
[53] Stephen Purpura, Victoria Schwanda, Kaiton Williams, William Stubler, and Phoebe Sengers. 2011. Fit4life: The Design of a Persuasive Technology Promoting Healthy Behavior and Ideal Weight. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*. Association for Computing Machinery, Vancouver, BC, Canada, 423–432. https://doi.org/10.1145/1978942.1979003
[54] Yizhou Qian and James Lehman. 2017. Students’ Misconceptions and Other difficulties in Introductory Programming: A Literature Review. *ACM Transactions on Computing Education (TOCE)* 18, 1 (2017), 1–24.
[55] Steven R. Rick, Aaron Paul Goldberg, and Nadir Weibel. 2019. SleepBot: Encouraging Sleep Hygiene Using an Intelligent Chatbot. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion* (Marina del Ray, California) *(IUI ’19)*. Association for Computing Machinery, New York, NY, USA, 107–108. https://doi.org/10.1145/3308557.3308712
[56] Pooja Sankar, Jessica Gilmartin, and Melissa Sobel. 2015. An Examination of Belongingness and Confidence among Female Computer Science Students. *ACM SIGCAS Computers and Society* 45, 2 (July 2015), 7–10. https://doi.org/10.1145/2809957.2809960
[57] Jill M. Simons. 2011. *A National Study of Student Early Alert Models at Four-Year Institutions of Higher Education*. Ed.D. Arkansas State University, United States – Arkansas. http://search.proquest.com/docview/910327556/abstract/97BE5A6AF7BB47B2PQ/1
[58] Dale R. Tampke. 2013. Developing, implementing, and assessing an early alert system. *Journal of College Student Retention: Research, Theory & Practice* 14, 4 (2013).
[59] Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869* (2015).
[60] Yi-Shun Wang, Ming-Cheng Wu, and Hsiu-Yuan Wang. 2009. Investigating the Determinants and Age and Gender Differences in the Acceptance of Mobile Learning. *British Journal of Educational Technology* 40, 1 (Jan. 2009), 92–118. https://doi.org/10.1111/j.1467-8535.2007.00809.x
[61] Claire E Weinstein, D Palmer, and Ann C Schulte. 1987. Learning And Study Strategies Inventory (LASSI). *Clearwater, FL: H & H Publishing* (1987).
[62] Jacqueline M Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. 2018. Measuring Young Children’s Long-Term Relationships with Social Robots. In *Proceedings of the 17th ACM conference on interaction design and children*. 207–218.
[63] Alex C. Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T. Iqbal, and Jaime Teevan. 2018. Supporting Workplace Detachment and Reattachment with Conversational Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173662
[64] Barry J Zimmerman. 1990. Self-Regulating Academic Learning and Achievement: The Emergence of a Social cognitive Perspective. *Educational psychology review* 2, 2 (1990), 173–201.
[65] K Zvereva, V Deviatkov, E Smirnova, and E Manyashev. 2020. Method Of The Student’s Motivation Assessment Using Smart Chatbot. In *INTED2020 Proceedings*. IATED, Valencia,Spain, 627–633.

A APPENDIX

A.1 Design Inquiry Survey

Q1 Your gender:
Male
Female
Other (please Specify) ____________

Q2 Your major
Computer Science
Other (please specify) ___________

Q3 Year of school:
Freshman
Sophomore
Junior
Senior
None of above

Q4 Please select all instant messaging apps you regularly use on your desktop:
Facebook Messenger
Whatsapp
Wechat
Slack
Skype
Telegram

Q5 Please select all instant messaging apps you regularly use on your phone:
Facebook Messenger
Whatsapp
Wechat
Slack
Skype
Telegram

Q6 Have you ever used Slack?
Yes
Maybe
No

Q7 (display if Q6 != No) How often have you used slack (opened the application) in the past month?
Daily
4-6 times a week
1-3 times a week
Rarely
Never

Q8 (display if Q6 != No) How often have you sent or replied to slack messages in the past month?
Daily
4-6 times a week
1-3 times a week
Rarely
Never

Q9 How familiar are you with using chatbots or conversational agents?
Extremely familiar
Very familiar
Moderately familiar
Slightly familiar
Not familiar at all

Q10 How interested would you be in a chatbot that helps you with your course-related activities (e.g. studying, homework)?
Extremely interested
Very interested
Moderately interested
Slightly interested
Not interested at all

Q11 Please rate the extent to which you agree/disagree with the following statements:
1 strongly agree
2 agree
3 neither agree nor disagree
4 disagree
5 strongly disagree

1) I'm confident that I can do an excellent job on my CS tests.
2) I'm certain I can understand the most difficult material presented in CS textbooks.
3) I'm confident I can understand the most difficult material presented by my CS programming instructor.
4) I'm confident I can do an excellent job on my CS assignments.
5) I am certain I can master the skills being taught in my CS class.
6) I believe I will receive an excellent grade in my CS class.
7) I'm confident I can learn the basic concepts taught in my CS class.
8) I expect to do well in my CS class.

Q12 Please rate the extent to which you agree/disagree with the following statements:

1) I have a difficulty getting started on a course project.
2) I have difficulty managing my time for a given assignment.
3) I would like to learn more insider tips from senior CS students and graduates.
4) I am familiar with the resources at the university that could help me with school work.
5) I know that I can get help from various school resources when I am in trouble.

Q13 Please indicate your familiarity about the following resources:
1 extremely familiar
2 very familiar
3 moderately familiar
4 slightly familiar
5 not familiar at all

1) Computer Science Resource Center
2) Advising Center
3) Teaching Assistant Office Hours
4) Course Instructor Office Hours
5) Stack Overflow
6) Online courses: Codecademy, MOOCs
7) Classmates who have previous taken the course

Q14 How useful would you find the following features in an application or chatbot?
1 extremely useful
2 useful
3 neutral
4 useless
5 extremely useless

1) Help in breaking up large tasks into smaller, more manageable goals
2) Assistance in managing my time for assignments and academic tasks
3) Reminders for completing my academic tasks on time
4) Computer Science tips from experienced and graduated students
5) Recommendations for academic resources available to me as a student
6) Feedback for how well I am managing my time and practicing good study habits
7) Connecting me to a tutor or advisor when I really need help

Q15 Is there any other help or support you feel you are currently not receiving that would help you with your course work? _______________

A.2 Design flowchart

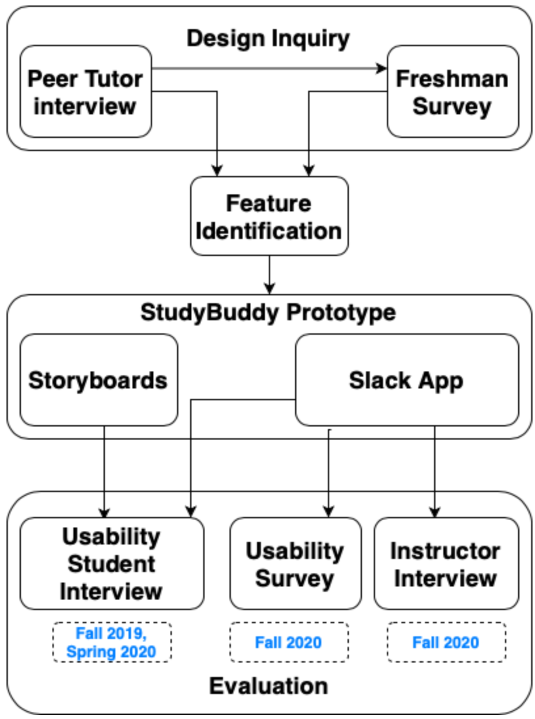

Fig. 5. Flowchart demonstrating the procedure of the study

A.3 Storyboards showed to participants during the usability evaluation

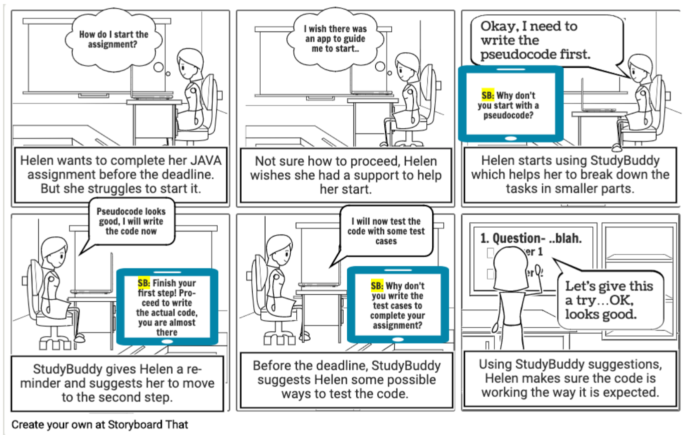

Fig. 6. First Storyboard related to task breakdown feature

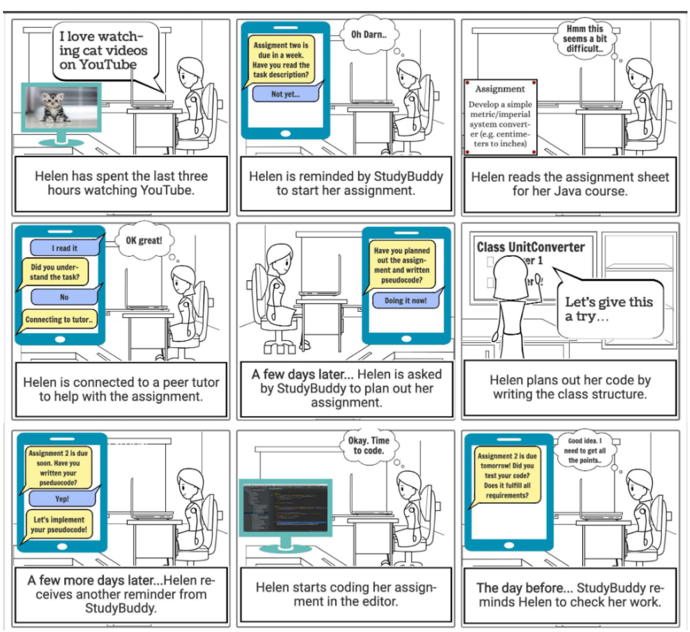

Fig. 7. Second Storyboard related to task break-down, scheduling and reminders and connecting to tutors features

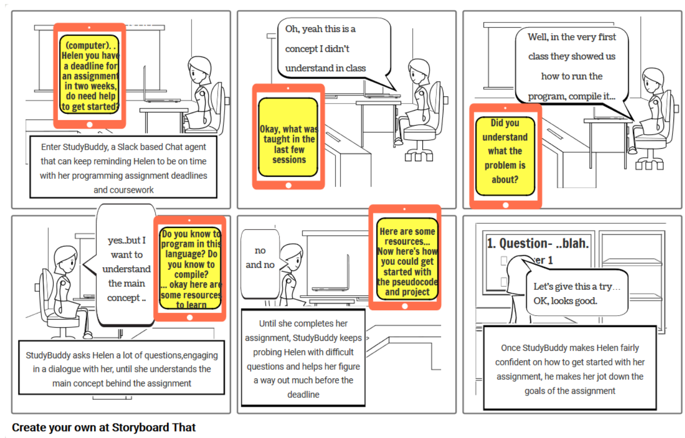

Fig. 8. Third Storyboard related to task break-down, scheduling and reminders and recommending academic resources features

A.4 Usability Survey

Q1 Are you 18 years or older?
Yes
No

UTAUT Questions
Q2 Regarding your normal learning activity, please indicate the degree to which you agree or disagree with the following statements. (1= Strongly disagree, 5= Strongly agree)

UTAUT - Self-management of learning:
When it comes to learning and studying, I am a self-directed person.
In my studies, I am self-disciplined and find it easy to set aside reading and homework time.
I am able to manage my study time effectively and easily complete assignments on time.
In my studies, I set goals and have a high degree of initiative.

Video page
Please watch this 3-minutes video demonstration of the designed chatbot functionalities.

Q3 Based on your impressions from the video demonstration of *Studybuddy*, please indicate the degree to which you agree or disagree with the following statements. (1= Strongly disagree, 5= Strongly agree) [items were mix-ordered when presented to participants]

UTAUT - Trust:
I believe *Studybuddy* would provide correct and useful information to improve study behaviors and practices.
Studybuddy could help improve how I study.
I believe *Studybuddy* could be trusted with study behaviors and preferences.

UTAUT - Effort:
The amount of effort to incorporate *Studybuddy* into study routines is acceptable to me.
Learning to use *Studybuddy* appears easy to accomplish.
I would find it easy to get *Studybuddy* to do what I want it to do.
Interaction with *Studybuddy* is clear and understandable.

UTAUT - Performance expectancy:
Studybuddy could help me complete tasks on time.
Studybuddy could improve my learning performance.
Studybuddy could help me achieve goals that are important to me.
Studybuddy could enhance my effectiveness of learning.

UTAUT - Satisfaction:
I felt delighted with *Studybuddy*.
I was very content with *Studybuddy*.
I was very pleased with *Studybuddy*.

UTAUT - Behavior intention:
If I had access to *Studybuddy*, I would use it.

If I had access to *Studybuddy*, I predict that I would use.
I plan to use *Studybuddy* in the future.

Design Questions
Q4 Based on your impressions of the video demonstration of *Studybuddy*, please rank each feature based on helpfulness to you:
Self-test and feedback for how well I am practicing good study habits (1)
Breaking down my large project into smaller, manageable tasks (2)
Scheduler and reminder for completing my assignments and projects on time (3)
Recommending academic resources (4)
Sending insider tips from experienced and graduated students (5)
Connecting to a human tutor when I really need help (6)

Q5 Why do you rank the most helpful feature in the above question? (open-ended)

Q6 Why do you rank the least helpful feature in the above question? (open-ended)

Q7 Regarding the 1st feature of : Survey that helps diagnose study habits. How often would you like to receive the survey prompt from the chatbot?
Weekly
Biweekly
Monthly
2-3 times per semester
Semesterly
Yearly

Q8 Regarding the 1st feature of : Survey that helps diagnose study habits. Would you like to share your survey results to others?
Yes, I want to have this option
No. I rather keep it only available for myself

Q9 If you are willing to share this result to others, to receive further help, whom would you like to share your study habits result with? (choose all that apply)
School / department
My academic advisor
My senior friends
My peer friends
My family
Others, please specify

Q10 What is your biggest concern about using *Studybuddy*? (open-ended)

Q11 If there is a chatbot (not necessarily *Studybuddy*) helping on my learning and my study habits, I worry about... (1-Not at all important, 2-Slightly important, 3- Neutral, 4-Important, 5-Very important, 6-Not sure [excluded from analysis])
The dialogue seemed too dull (1)
It seemed hard to communicate my needs to the chatbot (2)
The dialogue seemed too excessive before the chatbot can really help me out (3)

The bot doesn't personalize my needs (e.g., send more reminders if I tend to procrastinate) (4)
The bot isn't aware of my learning context (e.g., when assignment is due) (5)
The bot isn't adaptive as my study habits might change over time (6)

General Information and Demographics
Q12 What is your current major?
Computer Science
Information Science
Other (please specify)

Q13 What is your current year of study?
Freshman
Sophmore
Junior
Senior
Other (please specify)

Q14 What is your gender?
Male
Female
Nonbinary

Q15 (display if Q13 = Freshman) What prior CS experience do you have? (choose all that apply)
High school classes
AP exam or class
Online courses
Experience from another college-level source
Professional experience
Self-taught

Q16 How many CS courses have you completed?
0-1
2-3
4-5
6+

Please leave your email address if you would like to participate in being randomly chosen to get $25 cash upon the completion of the study. We will contact you directly if you were selected. (You can skip this question if you choose opting out)

Received June 2020; revised October 2020; accepted December 2020