Modeling and Supporting Middle School Mathematics Collaboration using Student Motivation across Different Digital Learning Platforms

by

Ishrat Ahmed

B.Sc., Bangladesh University of Engineering and Technology, 2014

Submitted to the Graduate Faculty of

the Department of Computer Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF COMPUTER SCIENCE

This dissertation was presented

by

Ishrat Ahmed

It was defended on

July 17, 2023

and approved by

Dr. Erin Walker, School of Computing and Information, University of Pittsburgh

Dr. Diane Litman, Department of Computer Science, University of Pittsburgh

Dr. Jacob Biehl, School of Computing and Information, University of Pittsburgh

Dr. Yoav Bergner, Learning Sciences and Educational Technology, New York University

Modeling and Supporting Middle School Mathematics Collaboration using Student Motivation across Different Digital Learning Platforms

Ishrat Ahmed, PhD

University of Pittsburgh, 2023

Adaptive collaborative learning support (ACLS) aims to facilitate collaborative activities by providing intelligent feedback and support based on students' collaboration. Existing ACLS systems have been applied in various collaborative environments, ranging from co-located collaboration in the classroom to online learning environments. While these technologies show promise, they primarily focus on supporting students within a single activity in a given platform and do not consider that students often collaborate across multiple learning platforms. Furthermore, the adaptive supports are primarily based on student participation and do not take into account how students are motivated to collaborate. Student motivation is critical during collaboration as it can encourage active participation and enhance overall learning outcomes. Motivation can influence how a student chooses to collaborate when participating in different types of learning platforms, along with the affordances of the platforms. Therefore, students' motivation may interact with different platforms' affordances, requiring individualized collaborative support. Hence, in addition to the cognitive model within an ACLS, the adaptive support model should also consider student motivation as it interacts with context. To address this gap, we have developed UbiCoS (Ubiquitous Collaboration Support), a collaborative learning technology that encompasses three different digital platforms: a discussion-based synchronous environment, a question-answer-based asynchronous environment, and a virtual teachable agent. The specific collaborative skill we are interested in is help-giving, as giving help to others encourages students to reorganize and clarify content, reflect on misconceptions, and fill gaps in their knowledge.

We deployed UbiCoS in a middle school classroom, where several experiments were conducted to support help-giving across the different platforms. We investigated how the same students collaborated on the three platforms and identified factors that influenced their help-giving. Using the collected data, we developed an explanatory participation model where individual characteristics and platform properties intersect, which can then be used to design collaborative support. Additionally, we discovered that platform features affect individual characteristics, leading to changes in motivation when collaborating on different platforms. To assess this dynamic aspect of motivation, we developed an interactive persona tool for students to report their motivation prior to each digital collaborative activity. As a final step, we designed collaborative support that considered student motivation and implemented a badge system to reward student participation. The thesis makes significant contributions in the fields of Human-Computer Interaction, Computer Science, and Learning Science in the following areas: 1) The design of a collaborative learning environment encompassing multiple platforms instead of a self-contained environment, 2) The adaptation of traditional Intelligent Tutoring Systems (ITS) to model student collaboration using individual and platform characteristics, 3) The design of an interactive tool for the dynamic assessment of student motivation, and 4) The provision of personalized support across platforms, using student motivation to facilitate effective collaboration.

Table of Contents

List of Tables

List of Figures

Preface

All praise to Allah SWT for giving me the patience to complete this journey despite numerous personal and professional challenges. Throughout this journey, I have been fortunate to collaborate with numerous individuals who have offered their guidance, support, and expertise. I would like to thank the following:

My advisor Erin Walker, for her invaluable mentorship and insightful feedback. Her guidance has been instrumental in shaping the direction of this research and pushing me to strive for excellence. She guided me to approach a research problem and trained me to become an independent researcher. She believed in me when I did not believe in myself. Not only in academic life, but her unwavering support when my baby passed away in 2021 is one of the reasons I am able to complete this journey today.

My dissertation committee members, Diane Litman, Jacob Biehl, and Yoav Bergner, for their expertise, thoughtful insights, and constructive criticism. Specially Yoav, who played a crucial role in refining the quality and rigor of the work.

My collaborators at Arizona State University and the University of Pittsburgh, Ruth Wylie and her students, Victor, Shang, and many others, who assisted with the classroom studies, helped with the analysis, and co-authored papers.

My colleagues in the FACET lab for providing a collaborative environment throughout this journey. Among them, now a dear friend, Deniz's encouragement, insightful discussions, and positive energy have been a constant source of inspiration for me.

My family, including my parents, my husband Masudul, and my brother, my in-laws, for their unconditional support, love, and understanding. My angel son Ishmaam, for teaching me resilience, and my earthside daughter Nusaibah, for teaching me determination.

Last but not least, all my friends who were patient enough to listen to my endless rants, doubts, and moments of despair. Your friendship has been a vital source of balance in my life, reminding me of the importance of taking breaks, enjoying moments of laughter, and cherishing the bonds we share.

1.0 Introduction

Intelligent Tutoring Systems (ITSs) are personalized learning technologies designed to monitor students' learning activities and provide individualized support when needed. A variation of such systems extended to collaborative activities is known as adaptive collaborative learning support (ACLS). It aims to design efficacious support that models students' collaboration and provides intelligent support to facilitate collaborative learning [115, 3, 81]. Technology-supported collaborative learning can enhance peer interaction, and both technology and collaboration can facilitate sharing and distributing knowledge [67]. For example, during collaboration, students are prompted to verbalize their thoughts, work collaboratively, ask questions, explain and justify their opinions, and elaborate and reflect on their knowledge [120, 22]. However, collaboration does not happen automatically [25] as students often struggle to collaborate. One possibility could be that the students have limited knowledge of interacting with each other when working in groups [62, 81]. As a result, students may fail to take complete advantage of the collaborative activities and need support to engage in productive interactions such as deep elaboration or constructing adequate arguments. Adaptive collaborative learning support systems are designed to structure collaboration so that particular promotive interactions emerge. Many ACLS systems have been applied in different collaborative environments, ranging from co-located digital collaboration in the classroom [11, 63, 115, 50] to online learning environments [46, 3, 48]. For instance, Harsley et al. [50] describe a system that supports pair collaboration in the classroom as the pair jointly engages in solving linked lists data structure problems. The system provides collaborative support utilizing a graphical interface, including pie charts and bar graphs to display group participation. In contrast, Haq and his colleagues [48] developed a framework supporting small group collaboration in an online platform. The system facilitated discussions around object-oriented programming questions, where students responded individually first and then had access to peer responses for reflection. The system provided feedback by analyzing group responses and selecting the most authentic answer as the collaborative response.

While these technologies show promise, they focus on supporting students within a single activity in a given platform and do not consider that students often collaborate across multiple learning platforms. The benefit of using multiple learning platforms is that the students can have a diversified learning experience since each platform offers unique features for collaboration. Collaboration in such platforms can either be synchronous (occurring in real-time and requiring the simultaneous participation of students) or asynchronous (occurring in delayed time and does not require the simultaneous participation of students). The synchronous mode of collaboration can be spoken or text-based, while the asynchronous mode of collaboration usually involves text-based threaded discussion [59]. Furthermore, these modes of collaboration can be in a public forum with geographically distributed students or in a private classroom with peers. With multiple platforms, students may prefer to participate in one platform compared to another. Moreover, as students learn how to construct knowledge on one platform collaboratively, they are expected to transfer these skills to another platform to solve similar problems. Facilitating the transfer of skills using ACLS might ultimately enhance students' collaborative learning abilities beyond a single platform.

To design ACLS for multiple platforms, it is crucial to understand how the student interacts across different learning environments, ultimately allowing a seamless and integrated collaborative learning experience for students. Students might behave differently on different platforms, and focusing interaction within a single context limits the potential effectiveness of the ACLS. Student behavior in a synchronous collaborative learning environment might be different than in an asynchronous collaborative learning environment [28, 84]. For instance, Davidson et al. [28] found that graduate students exhibited more responsive and reactive statements in synchronous chats compared to an asynchronous threaded discussion environment. In contrast, Oztok et al. [84] reported that threaded discussions fostered more complex and reflective statements than synchronous chats. They explored graduate students' social and cognitive presence in a synchronous and asynchronous tool within the same online environment. The results indicated synchronous messages contained more social and emotional content than asynchronous notes, and asynchronous notes had more cognitive processing words than synchronous messages. The contrasting behavior exhibited by students emphasizes the importance of exploring student interactions across various learning platforms to design effective collaborative support.

Traditional adaptive collaborative supports are typically based on monitoring students' progress and participation within the system [92, 63, 53, 3, 114, 50]. However, an additional area where these systems can be further extended is by considering student motivation when designing collaborative support. Student motivation is critical during collaboration as it can encourage active participation and enhance overall learning outcomes. Motivation can further influence how a student chooses to collaborate when participating in different types of learning platforms, depending on the affordances of the platforms. A student may have high domain knowledge but may need to be more motivated to collaborate with strangers in an asynchronous threaded discussion. So, students' motivation possibly interacts with different platforms' affordances and who they are collaborating with, requiring individualized collaborative support for students on different types of platforms. Hence, in addition to the cognitive model within an ACLS, the adaptive support model should also consider student motivation as it interacts with context. In this dissertation, we take a step toward developing a new learning technology that facilitates student collaboration across multiple learning environments. Our goal is to design adaptive support that leverages student motivation to enhance the quality of collaborative discussions.

In this work, we created a suite of computer-supported collaborative learning (CSCL) environments called UbiCoS (Ubiquitous Collaboration Support) encompassing three different platforms: a discussion-based synchronous platform, a question-answer-based asynchronous platform, and a virtual teachable agent. It was designed and implemented by an interdisciplinary team including computer scientists, learning scientists, middle school students, and middle school teachers [5]. UbiCoS aims to improve students' collaborative behaviors across multiple activities and platforms. The particular collaborative skill we are interested in is help-giving, which is an activity where students interact with their peers, explain to one another, and provide feedback and examples [58, 120]. In addition, effective collaboration also requires that students build on, relate to, and refers to their collaborators' work, referred to as 'Transactivity' [14]. This allows the students to elaborate or answer their partners' questions and modify their own explanations, which leads to the co-construction of knowledge [55]. Hence, transactivity is an important part of successful help-giving. We focus on students' help-giving in a middle school mathematics class where students have opportunities to answer questions posed by their peers and provide feedback on their work. In UbiCoS, students give help across four contexts: face-to-face small group discussions, digital discussions with peers within a digital textbook called ModelBook, question and answering using Khan Academy, and tutoring a teachable agent. The same students collaborate on these platforms to solve open-ended problems. Our first overarching research question is to investigate **RQ1: How does student collaboration vary across the three different learning platforms?** By gaining insights into the variations in student help-giving, we can design personalized and adaptive support systems that cater to students' preferences and each platform's unique characteristics and affordances.

As mentioned before, student motivation might be a key factor influencing collaboration in UbiCoS. We used student motivation as a pivotal factor in developing different aspects of this cross-context collaborative support. First, the modeling of student collaboration is an important step in ACLS systems. While many modeling approaches have been used to support collaboration in ACLS systems, they often focus on student participation and disregard factors like motivation and platform affordances. Yet, collaboration can vary depending on students' motivation and preference to collaborate on different platforms. For instance, students may be willing to collaborate synchronously with their peers but hesitate to collaborate in a public question-answer-based forum. The key question is, therefore: **RQ2: How can we design an explanatory model using student motivation and contextual factors to explain students' collaborative behavior?** In UbiCoS, students collaborate in three different learning environments, and by modeling such collaboration, we seek to better understand student participation in these platforms. Second, to effectively model motivation to provide support, we need to assess student motivation related to collaboration in mathematics. One approach to investigating motivation in learning focuses on its dynamic and responsive nature, considering contextual factors and their influence on the learner-environment relationship [82]. Thus, following this approach, we focus on **RQ3: How can we assess student motivation dynamically and within context?** We developed an interactive tool that is contextually embedded in the digital textbook that students used to report their motivation at the beginning of each digital collaborative activity. Finally, to facilitate collaboration, we explored **RQ4: How can we design adaptive collabo-**

rative support using student motivation and context? To answer this, we designed collaborative support leveraging the help-giving theory [120] and used student motivation and platforms. The students' self-reported motivation data was used to display support and reward students' participation through badges, aiming to facilitate collaboration across the different platforms.

To summarize, this dissertation investigated the broad areas of ACLS development: cross-context collaboration technology, modeling student participation using motivation and context, assessing motivation within context, and designing support using motivation and context. Table 1 describes the overarching research questions and associated design elements, methods followed, and sub-research questions for each research question.

1.1 Research Contribution

This dissertation contributes to Computer Science and Human-Computer Interaction (HCI) by introducing innovative approaches to education research and learning science. We developed UbiCoS, a cross-context collaborative learning environment that enables a system that can function in the background across multiple learning platforms, encouraging new means of fostering and assessing collaboration.

For the Computer Science community, we developed a theorized, explanatory model for student collaboration in the realm of small data using student motivation and platform characteristics. Although there is room for improvement, we believe the process of generating this model is unique and generalizable to other study contexts.

For the HCI community, we developed an interactive tool for assessing motivation within context. In many works of literature, motivation is considered to be dynamic and changes within the learning steps. Following the dynamic aspect of student motivation, we used the tool to measure motivation prior to each collaborative activity and provide relevant support.

For the Learning Science community, this research contributes to collaboration theory by exploring student interactions across platforms and offers practical recommendations for implementing these technologies in the classroom.

Table 1: Overarching and respective sub-research questions in the four major areas of ACLS development. Design and method indicate the product developed and the process followed to address each overarching research question. "Ch#" stands for chapter number.

Area	Category	Description
Technology	RQ1	How does student participation vary across the three different learning environments?
	RQ1 Design	UbiCoS System
	RQ1 Method	Design Based Research
	RQ1a	How do individual students' help-giving interactions vary across Modelbook and Khan Academy? (Ch#3)
	RQ1b	How does student help-giving interaction across different platforms predict learning and motivation? (Ch#3)
	RQ1c	How do individual differences (motivation and prior domain knowledge) predict student help-giving behavior across multiple platforms? (Ch#3)
	RQ1d	How do the interaction quantity and quality differ across the three platforms in the three cycles? (Ch#3)
Modeling	RQ2	How can we design an explanatory model using student motivation and contextual factors together and explaining students' collaborative behavior?
	RQ2 Design	Explanatory Modeling Approach
	RQ2 Method	Logistic Regression
	RQ2a	Which factors motivate or inhibit students help-giving behavior on different platforms? (Ch#4)
	RQ2b	How do we model students' constructive participation using individual and platform characteristics to explain their participation? (Ch#4)
	RQ2c	How well does the computational model explain the students' constructive participation in unseen data? (Ch#7)
	RQ2d	Is the computational model transferrable beyond the context it was built?(Ch#7)
Assessment	RQ3	How can we assess student motivation dynamically and within context?
	RQ3 Design	Contextually Embedded Interactive Tool
	RQ3 Method	Co-design sessions & Design Study & Classroom Study
	RQ3a	How did the students use the interactive persona tool? (Ch#5, Ch#6)
	RQ3b	How does interaction with the persona tool relate to students' collaborative behavior across the different platforms? (Ch#6)
Support	RQ4	How can we design adaptive collaborative support using student motivation and context?
	RQ4 Design	Collaborative Support Design using Sentence Starters and Badges
	RQ4 Method	Design Study & Classroom Study
	RQ4a	How did the students use the support (sentence starters and badges) across the three platforms? (Ch#6)
	RQ4b	What is the relationship between student support usage and student participation? (Ch#6)

1.2 Outline

In this chapter, I introduced the motivation, the overarching research questions, and the research contribution of the thesis. The rest of the thesis is organized as follows: Chapter 2 reports related work on the help-giving skill, adaptive collaborative support systems, and modeling and assessing dynamic motivation. In Chapter 3, I describe the collaborative learning environment, UbiCoS; the iterative cycles of the design-based research study, and present the quantity and quality of student help-giving participation across the three platforms in the three cycles. Chapter 4 illustrates the results of a semi-structured interview where we extracted factors influencing student help-giving participation across the three platforms. I then describe how we used a subset of those factors and the platform characteristics to develop a computational model to explain student participation. I present the results of the model. Chapter 5 describes a design study where we implemented the initial version of an interactive tool for assessing motivation within the context and the primary adaptive support. Chapter 6 describes a classroom study where we revised the interactive tool and the adaptive support based on the observation from the design study. We investigated and reported how students used the interactive tool to report their motivation and how the students used the adaptive system. Chapter 7 describes the generalizability and transferability of the computational model (developed in Chapter 4) across different study contexts. Finally, I end with Chapter 8, summarizing my work, acknowledging the limitations, and presenting future research avenues.

2.0 Related Works

This chapter centers around exploring work related to adaptive collaborative learning systems. We begin by examining the specific collaborative skill, help-giving, and how researchers facilitate this behavior. We proceed by describing existing collaborative intelligent tutoring systems and their feedback mechanisms. Additionally, we summarize the approaches employed by intelligent tutoring systems to assess the dynamic aspect of motivation. Furthermore, we discuss the utilization of gamification elements, such as badges, within the domain of the intelligent system and explore their potential extension for ill-defined collaborative systems.

2.1 A collaborative skill: Help-Giving

There are different methods to implement collaborative learning, such as cooperative small group learning [57, 102], reciprocal teaching [86], and classwide peer tutoring [42]. Peer-directed small-group learning is a topic of much interest to educational researchers. The benefits of such an approach have been highlighted in a number of empirical studies, e.g., Slavin et al. [101] reported how cooperative small group problem-solving is used to improve student learning and student motivation; Webb et al. [119] investigated different kinds of peer interaction that influence student learning in small groups; and Fuchs et al. [43] taught students to provide conceptual explanations (i.e., reasoning and explanation) in a peer-mediated group learning environment. Most of these peer interactions included high-level elaboration (i.e., explanations) and low-level elaboration (i.e., answer to a problem, procedural information, managerial information) [119]. These researches focus on a key aspect of peer interaction, i.e., students' help-giving behaviors during group learning [121] where students interact with their peers, give explanations to one another, and provide feedback and examples [58, 120]. Help-giving is one of the promotive interactions for collaboration [58], and has cognitive benefits as giving help encourages students to reorganize and clarify content, reflect on the misconceptions, and fill gaps in their knowledge [120]. Webb and colleagues have thoroughly investigated the relative impact of interactions during in-person collaborative tasks and concluded that giving and receiving elaborated explanations lead to greater learning gains [120]. Different studies [119, 79] further explored different factors influencing students' help-giving behavior. According to Webb [119], factors such as the characteristics of the individual (e.g., ability, personality, gender), the group's composition, and the type of questions asked predicted the level of elaboration in response to students' questions and errors. Nattiv et al. [79] conducted a similar analysis to investigate whether the helping behavior and achievement gain generalize to elementary grades as opposed to Webb's approach (i.e., seventh and eighth grade). They found that high achievers gave more explanations, and low achievers often asked for and received explanations.

While Webb and others focus on elaboration and its relation to learning, this approach misses the socio-constructive element of peer-directed small-group learning. This element refers to an individual's learning that occurs because of their interactions in a group [85]. The learning occurs when students build on each other's ideas while giving and receiving help, sharing knowledge, and resolving conflicts between their own perspectives and those of other students who are engaged in a collaborative process of knowledge building - another key piece of help-giving. The process of responding to each other's comments is known as transactivity [14, 105], a construct well-researched in the domain of educational psychology and computer-supported collaborative learning. Providing high-level help and engaging in transactive behavior are critical components of the knowledge-construction process during small-group peer interaction. In this process, the learners' mental functioning develops as they internalize and transform the content of social interaction [113]. One student's comment acts as a source for the other student's reasoning and knowledge construction. A lot of research on collaborative dialogs focuses on transactivity [94]. Thus, in this thesis, I focus both on elaboration and transactivity.

Previous works related to help-giving have trained the students to give high-level elaboration prior to the collaboration [121, 43, 79]. They primarily concentrated on face-to-face collaborative environments. Recognizing its advantages, several intelligent tutoring systems have also explored the concept of help-giving as a collaborative skill [115, 3]. Walker et al.

[115] investigated help-giving in a peer tutoring context where students worked in pairs, and Adamson et al. [3] investigated transactivity in a synchronous chat environment to provide adaptive support. Traditionally, efforts to facilitate help-giving among peers or within small groups have focused on improving the quality of help, emphasizing the students' cognitive processes without explicitly targeting the motivational factors that influence help-giving. In this thesis, our aim is to support students in giving help across various learning platforms by utilizing students' motivation associated with help-giving.

2.2 Systems with Adaptive Support for Collaboration

In the previous section, we described a specific type of collaborative activity, help-giving, and its benefits. It is evident that any kind of collaboration doesn't happen spontaneously [25] and requires much effort. Research continues to determine ways to support collaboration to foster productive interaction, leading to desired domain learning and collaborative skills. Existing research suggests that to support individual needs during collaboration, adaptive solutions are more effective than traditional static support. These adaptive systems can either be systems designed solely for facilitating collaboration, or they can be intelligent tutoring systems that were extended to support student collaboration.

Numerous systems are built solely with the purpose of providing adaptive support to student collaboration [111, 29, 92, 63, 53, 61, 3, 108]. Vizcaino et al. [111] developed HabiPro, a collaborative software designed to support good programming habits among first-year undergraduate students. Depending on how students collaboratively (groups of two or more) solve a problem in a shared window and how they use one of the four help buttons, the system provides feedback on both social and task-oriented aspects of group learning. Constantino-González and Suthers [29] describe COLER, a web-based collaborative learning environment in which undergraduate students can solve database-modeling problems using a shared workspace and communicate via a chat window. The students collaborate synchronously in small groups at a distance. LeCS, another web-based collaborative case study system [92], provides domain- and collaboration-based feedback. Kumar et al. [63] describe an ACLS, CycleTalk, where students worked in pairs in a simulation-based learning environment. They reported that the support implemented with tutorial dialogue agents is better than static support at increasing domain learning. GroupLeader [53] assists students working remotely in collaborative groups. This system analyzed the log files for each student and for each group and assessed the group's collaboration level based on the following three parameters: promotive interaction, social skills and group processing, positive interdependence and individual/group accountability. Karakostas and Demetriadis [61] demonstrated that students' learning improved when collaborating students received adaptive domain support in the form of prompts relating to important concepts the students had missed in their discussion rather than a static resource that described all important domain concepts. Adamson et al. [3] used conversational agents for scaffolding online collaborative learning discussions that would encourage students to articulate and elaborate their own lines of reasoning and challenge and extend their teammates' reasoning. Tsan [108] developed two collaborative support features in a block-based programming environment to support upper-elementary students' collaborative dialogue and reduce the number of conflicts they encounter.

In contrast to the above systems, much of the research extended ITS focused on individual learners to support collaboration. These systems focused on merging the affordances of both ITS and CSCL to capitalize on the benefits of group learning and adaptive support [49, 36, 11, 34, 81, 50, 48]. Harrer et al. [49] investigated whether a cognitive tutoring framework can support collaboration while students use collaborative software. Students collaborated in a shared workspace performing a UML modeling task, which the system analyzed using network graphs. The system intervenes if it notices that a student is going down a path that leads to incorrect solutions. In addition, [49] proposed to give feedback related to task coordination (i.e., distributing the work among group members and spending an appropriate amount of time in different phases of the collaboration) by analyzing the logs obtained from the collaboration tool. Dragon et al. [36] designed Collaborative-Rashi, a coaching system to support collaboration, on top of an ITS for ill-defined domain spaces (geology, biology, art history, and forestry) where the students engage in inquiry-based learning. Systems such as COLLECT-UML adaptively support students in collaborating on the design of UML class diagrams and have been shown to lead to greater knowledge of collaboration over an individual learning system [11]. Diziol et al. [34] developed two collaborative extensions to an intelligent tutoring framework: first, they used adaptive scripts for collaborative problem-solving aimed to support student interaction. Students worked in a peer collaboration scenario where they worked together to solve an algebra problem. Adaptive interaction support was provided when the system detected student behaviors such as trial and error and hint abuse and provided feedback based on those behaviors. In the second extension, [34] enhanced the adaptive script to support reciprocal peer tutoring and give feedback to the peer tutors. The system compared the peer tutor's action (correct or incorrect) and displayed a social prompt and the cognitive tutoring support that an individual learner would have received. The system also provided hints on demand to the peer tutor to assist the tutors in elaborating on their existing knowledge and constructing new knowledge. Olsen and colleagues [81] extended a cognitive tutoring authoring tool and developed collaborative tutors focused on procedural and conceptual knowledge to support elementary students' collaborative behavior. The students collaborated to solve fraction problems using their own computers via a shared workspace and communicated via speech. The collaborative tutor provided the usual cognitive and social support using embedded scripts. The scripts supported collaboration through the distribution of responsibility to create accountability and independence. Harsley et al. [50] modified an existing ITS in structuring collaboration, called Collab-ChiQat Tutor, to support collaboration between pairs of students as they jointly engage with the system in pair programming. Haq and his colleagues [48] developed a framework supporting small group collaboration in an online platform. The system facilitated discussions around object-oriented programming questions, where students responded individually first and then had access to peer responses for reflection.

All of these adaptive support systems, irrespective of how they are built, may vary in how they assess student collaboration and how they provide feedback to the students. However, one common thing across all of these systems is that they focus on a single learning environment, and each has an adaptive support system dedicated to it. In this thesis, we primarily explore the design of a single adaptive system that supports collaboration across multiple learning platforms instead of separate intelligent systems for each platform.

2.2.1 Modeling/Assessing of Student Motivation within ITS

Up until now, all the adaptive systems described above provide personalized feedback based on student interactions by tailoring learning materials (e.g., displaying easy problems and moving to harder problems based on student performance). One key aspect of student participation missing/neglected here is student motivation. Motivation is one of the learner's characteristics that influence learning processes. It is a major driver of engagement both in traditional and online learning environments. In traditional classrooms, the teachers can monitor the students' motivational state and adapt the teaching strategy. In ACLS systems, in addition to the cognitive model, it should also include motivation in the learner model to adapt instructions accordingly. Much of the recent efforts are given in this area where intelligent tutoring systems include motivation and provide personalized support to individual learners [31, 91, 76, 56, 117, 78, 98, 82].

To model student participation effectively and adapt to a student's motivational state, one of the first things the ACLS systems need is to include the assessment of motivation in the ITSs learner model. The most common form of assessing motivation is through surveys or questionnaires. However, this approach has limitations in an ITS context. In most cases, the assessment is done at the beginning of the study, which misses two important aspects of motivation: 1) motivation is influenced by the environment [72], so it should be examined in the context of events, and 2) motivation is dynamic and changes over time [35], so it should be assessed as such. This aspect of motivation has resulted in the adoption of different techniques to assess motivation. For example, some ITSs address dynamic assessment using an interface embedded with questionnaires and sliders [31] or utilizing a single self-report question [78]. Vicente et al. [31] used a motivational slider and asked the students to self-report their motivational state while using the ITS called MOODS (MOtivational Diagnosis Study). The students were asked to report their motivational state (effort, confidence, sensory interest, cognitive interest, relevance, satisfaction) at any time during their interaction with the system. Similarly, the system used in [78] utilized a single self-report question after each activity to label a student as 'motivated' and 'unmotivated.' However, there are a few challenges with utilizing questionnaires to assess motivation. For example, there could be difficulty in understanding vocabulary typically used to assess motivation [40, 44], and survey fatigue which can lower the quality of the responses/response rates [90]. Also, administering such questionnaires during tutoring can be too intrusive [117] and disruptive to student participation [76].

The other method for assessing motivation is using physiological data (e.g., capturing heart rate and galvanic response, eye gaze, facial expressions) to model motivation-related behavior and provide intelligent support [91, 76, 98]. By combining evidence from learner's eye gaze and interface actions, Qu et al. [91] used a Bayesian model to infer a learner's focus of attention as an estimate of motivation in ITS (i.e., a learner's degree of confidence, confusion, and effort). The evaluation of the model revealed that the recognition accuracies are above 70% for the learner's motivation compared to a model using a human tutor's observation as a baseline and another model that used the learner's self-reports as a baseline. McQuiggan et al. [76] developed a dynamic model to capture students' motivation (self-efficacy) in ITS using Naive Bayes and decision tree models. The dynamic model learns from pre-test data, physiological data (heart rate and galvanic skin response) of students captured with a biofeedback apparatus, and interaction data of students in their learning environment. The results indicate that the dynamic self-efficacy model predicted students' self-efficacy more accurately than a static model (excluding physiological response data) generated from data obtained using a validated self-efficacy instrument. Santos et al. [98] explored the use of convolutional networks in detecting levels of intrinsic motivation (interest and joy) using visual cues from students' facial expressions. While these approaches used alone or combined with machine learning methods can provide an automatic and often accurate estimation of the student's motivation [82], they have notable challenges. They are often costly and thus difficult to scale as they require sensors, cameras, and other devices to capture a continuous measurement of physiological data. Moreover, learners might react negatively toward these sensors [78]. Also, it depends on the type of motivational construct, e.g., boredom and confusion can be extracted using such physiological data – but for the confidence levels, it might be challenging for the sensors to detect these constructs.

Analyzing the log files is another approach that has been used extensively to assess student motivation. Johns et al. [56] developed a dynamic mixture model to infer student motivation (unmotivated-hint, unmotivated-guess, and motivated) from a hidden Markov model (HMM) and student proficiency from an Item Response Theory (IRT) model. They used students' log data from an ITS to validate the model. Ward et al. [117] used Natural Language Processing methods (semantic cohesion measure) to measure student motivation (self-regulation, self-efficacy, and intrinsic value). They used the student utterance data from the ITSpoke tutor and hypothesized that there is a direct correlation between motivation and cohesion when using the tutor. They calculated the mean scores on the pre-and post-tutoring motivational measures and used them to calculate correlation with cohesion. However, they did not analyze the motivational constructs separately. While the machine learning models could be more sophisticated/accurate, it requires a reasonable amount of data and a set of features to train a model and accurately predict students' motivation state.

One notable challenge with the above-mentioned approaches, including utilizing physiological data or analyzing log files, is that it can leave the students without agency in reporting motivation, as the assessments are not directly captured by the students. Student agency encourages ownership of learning and engagement [4] that promotes academic performance and learning experiences [70]. The lack of agency can result in reduced engagement in learning activities, reducing confidence in their capabilities. Thus, by integrating student agency in assessing their motivation, we expect to attain a more accurate representation of student motivational state.

All these systems describe methods to assess student motivation and support individual learners. In UbiCoS, we aim to provide adaptive support for collaboration across multiple learning environments. Student motivation becomes more critical as motivation may vary not only throughout a learning episode [76] but it is likely the motivation may also vary across the platforms. So we are interested in the dynamic assessment that will capture students' motivation within the context. An ITS can then use this data to create personalized interventions encouraging participation. We aim to assess students' motivation dynamically and design for the agency to have students report their motivation within context. This approach will ultimately support us in adapting collaborative support in each context based on the reported motivation to encourage meaningful collaborative participation.

2.2.2 Designing Collaborative Support

In the intelligent tutoring systems domain, when to offer support is a widely debated topic. Adaptive support can be given prior to, during, and after collaboration. Prior to the collaboration, supports can take the form of scripts embedded in the interface [34, 81], prompts using sentence openers [92, 36] (where students complete open-ended sentences), hints using widgets (e.g., buttons) [111, 115]. Supports that are given during and after collaborations offer feedback assessing student interactions (assessment could be action-based [29] or dialog-oriented) to improve the quality of collaboration, eventually, student learning. This feedback can be given via automated conversational agents (can vary from spoken to text-based) [111, 29, 63, 3, 106], or displaying feedback on the graphical user interface [50].

ACLS systems vary greatly from each other as they use different methods to facilitate collaboration. Some systems facilitate dialogue between collaborating students by using sentence starters/openers and use that to determine what type of support to provide [92, 36, 53]. In LeCS [92], students use sentence openers to facilitate the process of reaching an agreement in the case discussion in a chat window. The students are asked to classify their own utterance, which is later compared against an ideal dialogue model to initiate an intervention. Collaborative-Rashi (or C-Rashi) [36] used sentence openers to provide support in a synchronous chat environment where students have solved problems in an ill-defined domain. The support was designed to facilitate an in-depth, on-topic discussion and to provide a coherent view of the argument. Groupleader [53] used 46 sentence openers to label their utterance and provide support. However, one issue related to sentence starters is that the students often do not consistently select sentence starters that match the statements' content, perhaps because they do not know how to use them effectively or are not motivated to do so, and hence support can be inaccurate [53, 65]. This has been recently improved using machine learning classifiers to automatically classify student contributions and trigger support as necessary [94, 75]. Another common way to help students collaborate is to embed adaptive scripts within the interface [34, 81]. Diziol et al. [34] provided adaptive interaction support via scripts for collaborative problem-solving. The system detected student behaviors such as trial and error and hint abuse and provided feedback based on those behaviors. They also enhanced the script to support reciprocal peer tutoring, where the system compared the peer tutor's action (correct or incorrect) and displayed a social prompt and cognitive tutoring support. The collaborative tutors developed by Olsen and colleagues [81] used embedded collaborative scripts to provide social support for students.

In order to give adaptive feedback, a system has to be able to automatically assess student collaboration. Different methods are used within ACLS systems to assess and compare student collaboration and then provide feedback to the students. The simplest way to assess collaboration and provide support is to use metrics other than the content of the dialogue itself, such as counting individual's utterances, and then encourage those who are not participating to participate more [92]. ACLS systems such as COLER [29] and [49] focus exclusively on student problem-solving actions in a shared workspace to give feedback, ignoring the content of language communication between collaborating students. In contrast, ACLS systems that analyze the content of student utterances may use simpler methods (e.g., sentence starters, keywords) or use advanced machine learning or natural language processing methods to categorize student participation and provide collaborative support [115]. Walker et al. [115], in their reciprocal peer tutoring context, used automatic classifications, student self-classifications, and domain information in the dialog to asses the actions peer tutors take during collaboration. The chat window had intelligent support to assist peer tutors in providing better help to the tutees. While all these studies used some sort of student interaction to provide feedback, Dennis et al. [33] reported adaptation of feedback (using conversational agents) based on learner characteristics and performance.

After assessing the students' participation in different ways, ACLS systems attempt to provide support in different ways to facilitate collaboration. For example, some systems use conversational agents to provide peer interaction support: HabiPro system [110] uses a simulated agent as a group member and intervenes in the following three cases as it may decrease the benefits of learning in collaboration: off-topic conversations, students with passive behavior, and problems related to students' learning. During the collaborative activity, the simulated peer uses the group model to compare the current interaction state and propose actions (e.g., suggests continuing with the problem in case of off-topic conversation, gives hints in case students do not have enough knowledge). In COLER, the author seeks to facilitate effective group-learning interactions by implementing a coach as a personal assistant to each student. The coach gives feedback when there are problems in the quality of solutions, differences in individual and group solutions, and differences in the learners' participation levels. [63] used a tutorial dialogue agent to provide interactive dynamic support within a synchronous collaborative chat environment. Both [106, 3] used conversational agents to support collaborative activities based on the Academically Productive Talk (APT) framework. This framework is consistent with the transactivity perspective. Tegos et al. [106] describe MentorChat, a teacher-configurable domain-independent conversational agent that uses natural language to scaffold peer discussions in synchronous (text-based) collaborative learning settings. The agent adaptively supports each group discussion by monitoring dialogue and identifying opportunities for intervention. Similarly, Adamson et al. [3] investigate the use of conversational agents to scaffold online collaborative learning discussions. Their system Bazaar provides collaboration support to distributed groups of learners collaborating synchronously. The system prompts students for explanations in the context of group discussion with the goal that students will articulate their own reasoning and evaluate and challenge others' reasoning. However, using such an agent-based approach has certain limitations. First, it is very content-specific, hence not scalable [3] across different platforms. Second, having one coach per student implies that students' behavior is more monitored, which may provoke students to collaborate more with their coach than their partners, producing the opposite effect of the coach intent [110]. In contrast to these conversational agents, non-conversational adaptive supports are also prevalent within the systems in order to facilitate collaborative dialogues and problem-solving. The other type of feedback given to the students is using visual tools. In Collab-ChiQat Tutor [50], the system used a graphical user interface to give collaboration feedback such as hints about successful collaboration, visual charts explaining peer collaboration (number of spoken utterances between partners, comparison of the number of compile errors vs. success per problem), and overall group score.

2.3 Providing Feedback using Badges within ITSs

In the previous section, we get a brief overview of how ACLS systems provide support and feedback to students. The systems gave feedback based on students' actions in a shared workspace or student utterances. In case the students are using synchronous messaging to collaborate with each other, a common approach to give feedback is a conversational agent that assesses student utterances or discussions and provides feedback to the students [63, 3]. While this approach has been useful in collaboration and learning, this can also be disruptive during synchronous collaboration, e.g., if the agent gives feedback to the whole group, it can break the conversation flow among the group members; if the agent gives feedback to a particular student in the group, the student might get distracted. Furthermore, we wanted a unified feedback system for all three platforms. In UbiCoS, the same students collaborate via synchronous messaging in an asynchronous question-answer platform and teach mathematics to a virtual teachable agent. To give feedback on student participation, we combined the approach of giving prompts before/during each collaborative activity and giving feedback to the students after their participation via badges.

A growing number of online learning environments are incorporating game-like elements (e.g., leaderboards, points, badges, etc.) as it has the potential to address challenges around student motivation and to positively impact learning [66, 18]. Among the game-like elements, badges are one of the common ones used in small-scale online courses [24], MOOCs [83], and more recently in ITSs [2, 41, 103]. Abramovich et al. [2] used badges within an intelligent tutoring system and investigated how educational badges affect learning motivation for middle school students. In this system, badges were awarded for skill mastery or for continued use of the system. Tahir et al. [103] added badges in SQL-Tutor, an intelligent tutoring system teaching problem-solving in SQL. They found that the more badges the students achieved, the more time they spent interacting with SQL-Tutor.

These studies have shown that badges can effectively incentivize students to complete specific tasks or increase participation. However, applying feedback mechanisms has challenges in ill-defined domains where the solutions to open-ended problems depend on reasoned arguments rather than a fixed number of formal steps. So, we adopted a positive feedback approach where we assess and reinforce the positive aspects of a student's utterance to guide student collaboration. Positive feedback is effective because it reduces uncertainty in the students' knowledge [77]. We used badges to give such feedback based on the quality of students' utterances rather than the number of tasks to be completed or the correctness of a task. It may encourage students to self-reflect, or make them aware of their participation, facilitating collaboration across the platforms.

2.4 Conclusion

This chapter provides a comprehensive overview of existing research in the field of adaptive collaborative learning support systems (ACLS). It includes various aspects, e.g., the help-giving skill, adaptive support systems, and the provision of feedback using badges in an ITS domain. The chapter describes how ACLS has been developed as either a stand-alone system or by extending existing ITS and has been applied in different collaborative environments, from co-located digital collaboration in classrooms to online learning platforms. By synthesizing previous studies, this chapter lays the foundation for the subsequent chapters that describe the design and development of cross-platform adaptive collaborative support systems.

3.0 Student Help-Giving Participation in Multiple Learning Environments

In this chapter, we first describe *UbiCoS*, the collaborative environment we have developed where students can practice help-giving in multiple learning environments with different collaborators. We aim to design a system to support help-giving and improve interaction quantity and quality across these platforms. Using the first overarching research question *RQ1: How does student participation vary across the three different learning environments?*, we hope to understand how students give help on different platforms. This will give us insights into their collaboration and whether students benefit from multiple platforms. To achieve this, we deployed UbiCoS in a middle school classroom following a design-based study (DBR) approach in three cycles. We investigated students' help-giving and iterated on the technology to better facilitate students' help-giving participation. In this chapter, we address the sub-research questions *RQ1a-RQ1d* as listed in Table 1 in Chapter 1.

This chapter is structured as follows: In Section 3.1, we describe *UbiCoS*, the collaborative learning environment encompassing three different platforms: Modelbook, Khan Academy, and Teachable Agent. Next, in Section 3.2, we describe the three cycles of the design-based research method sharing the major refinements across the cycles to improve students' help-giving behavior. Section 3.3 describes our initial analysis using research questions *RQ1a* to *RQ1c*, where we used the digital interaction data from Cycle 1. We measured the quality of students' help-giving by developing a coding scheme. We investigated the influence of students' motivation (i.e., self-efficacy), attitude towards math, and domain knowledge on students' help-giving. This work [5] was published at the Artificial Intelligent in Education Conference. Finally, we iterated on the coding scheme to improve assessing the quality of students' help-giving behavior developed in Section 3.3.2.1 and addressed *RQ4* where we analyzed the quality and quantity of students help-giving across the different platforms in the three different cycles. This work [73] was published at the International Conference of the Learning Sciences.

3.1 UbiCoS: System Description

To facilitate help-giving activities across various platforms, we focused on two main approaches: (1) Creating a curriculum to foster productive interactions and foster relationship-building in both face-to-face and digital settings, and (2) Developing technology to embody the curriculum and provide support for these activities. This section provides an overview of the curriculum and introduces the three platforms integrated into the *UbiCoS* collaborative learning environment.

3.1.1 Curriculum Design

We specifically designed our curriculum to promote productive collaborative interactions. To accomplish this, our team includes a co-designer teacher who has taught middle school for 13 years and was trained in modeling pedagogy. Modeling pedagogy is a method based on constructivism, where students engage in small-group, open-ended investigation of a concrete problem that provides the basis for developing a conceptual model [54]. They express their models on whiteboards, give feedback on other groups' whiteboards, and ultimately have a whole-class discussion to arrive at a set of principles related to the models. Modeling activities promote discussion since there are often multiple right answers and many correct learning paths and are similar to other learning pedagogies such as problem-based learning [52] and invention as preparation for future learning [99]. We chose the modeling curriculum because it encourages collaborative interactions, develops learners' sense of community, and engages students with math practices such as explaining and critiquing data interpretations. To develop the curriculum, we met with our co-designer teacher an average of once every two weeks for two semesters. We selected three topics aligned with common core and state standards for eighth-grade mathematics: ratios and proportions, volume and surface area, and linear functions.

3.1.2 Platform Description

We introduced three digital platforms where students could interact to complement the curriculum and the small group and whole-class discussions. We describe each of them in the following three sections:

3.1.2.1 Modelbook

The first platform is a digital textbook called ModelBook that we developed to contain curricular materials (e.g., question prompts and homework assignments). It allows the students to log their work (e.g., students could upload photos of their whiteboards) and enable them to interact digitally with their classmates at critical moments. Modelbook incorporates several components to facilitate student interaction. In ModelBook, students can see two windows: text on the left and one of several interactive tools on the right. One tool is the Gallery, where the students can upload work they completed in face-to-face groups and are able to evaluate, critique, and provide feedback to others through discussion (see Figure 1).

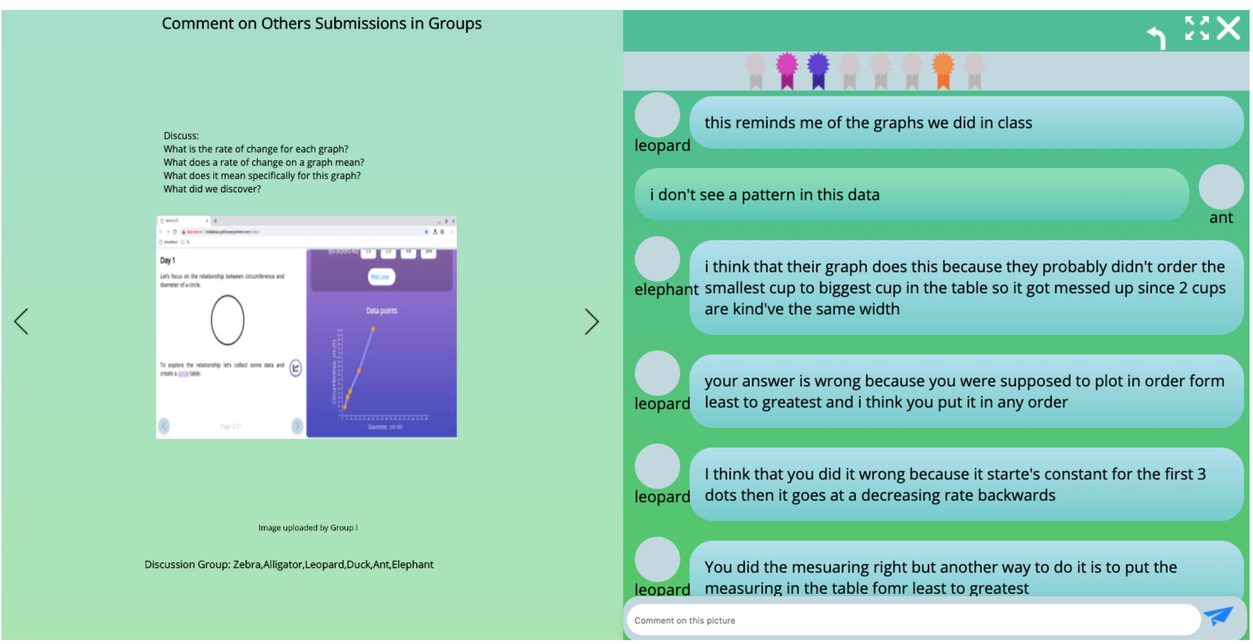


Figure 1: An example of the gallery discussion thread in Modelbook.

A second example is the Graphing Tool, where students input data points on a table and see the output on the graph. Finally, we have a place for a whole-class discussion called the Chat. With the teacher's guidance, students were encouraged to perform help-giving interactions in the gallery and general chat. ModelBook is designed to help students bridge their face-to-face and digital interactions. Students might work with one group to create a whiteboard, upload a photo of their whiteboard to a digital gallery, and then engage in a digital discussion about their whiteboard with another group.

ModelBook was built using the Django web framework. The application's front end is implemented using HTML, jQuery, and CSS. Templates within Django contain the static parts of the desired HTML output and provision for inserting dynamic content. For each tool in Modelbook, there are icons on the left hand of the application, which, when clicked, triggers a jQuery event to dynamically load the related interface on the application's right-hand side. All discussion threads were implemented using the "Pusher" service - a hosted API for quickly, easily, and securely adding a real-time bi-directional connection. We have used the default SQLite database that accompanies the Django framework. All user activity (e.g., uploaded images, messages) is stored in the database.

3.1.2.2 Khan Academy

The second digital platform is Khan Academy (www.khanacademy.org), a public question-answer platform where students watched short videos and participated in Q&A forums. While most known for its instructional videos, Khan Academy allows asynchronous collaboration with geographically distributed learners within the question-and-answer environment under each video. By answering questions, students articulated their understanding of concepts they learned and engaged in help-giving behavior with the broader public (see Figure 2).

For our curriculum, students posted responses to Khan Academy questions a few times during the study. To facilitate this activity, students were instructed to watch a related Khan Academy video (as homework or watch in class). They were asked to look for questions posted by other people and provide a response. Students' responses were then discussed in class the following day, and they were encouraged to post in class if they had not done their homework. An example of student participation in Khan Academy is shown in Figure 2.

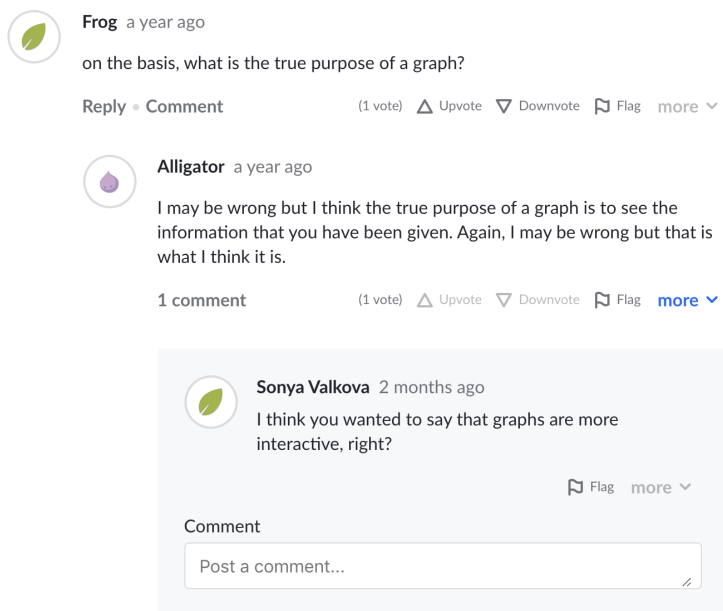


Figure 2: An example of discussion thread in Khan Academy.

3.1.2.3 Teachable Agent

The third digital platform is a Virtual Teachable Agent, a desktop version adapted from Lubold et al. [69]. In this system, students work individually with an agent to help them solve step-based mathematics problems. Students follow a worked example on an interface and explain each step to the agent. The agent responds in spoken dialogue. If the student provides the correct answer but fails to explain, the agent prompts them to elaborate (see Figure 3). Through these interactions, students practice help-giving skills in a context where domain knowledge is less of a factor since students are given the worked examples. Social anxiety is lessened since students are interacting with an agent versus a peer. During these interactions, we expect students to benefit by articulating their reasoning and responding to agent questions.

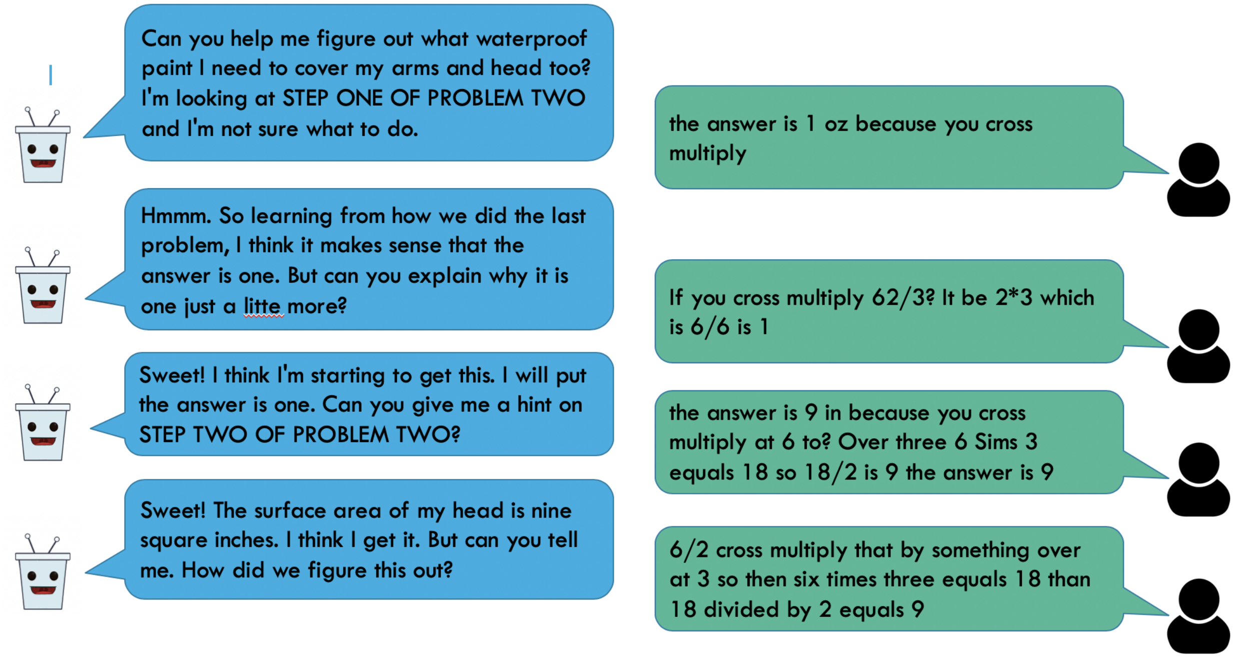


Figure 3: An example of the Teachable Agent interaction.

Each of the three platforms and face-to-face discussion represents a set of features that may influence how a student collaborates. For example, a student who is motivated to provide help because she wants to see her friends succeed may be more likely to contribute to a face-to-face discussion or Modelbook activity and less likely to answer a stranger's question on Khan Academy. A student with math anxiety and fears of making a mistake in front of a classmate may contribute less when interacting with a peer and may contribute more often on Khan Academy or with the Teachable Agent. The platforms also varied by interaction modality (speech versus text) and timing (synchronous and asynchronous). For this thesis, we focus on student digital interactions. Table 2 summarizes the different characteristics of the three platforms.

Table 2: Different characteristics of the three platforms.

Platforms	Modes of Instruction	Participants	Synchronicity	Public vs Private
Modelbook	Text, multiple-turn	Peers/Friends	Synchronous	Private
Khan Academy	Text, single-turn	Strangers	Asynchronous	Public
Teachable Agent	Speech	Agent	Synchronous	Private

3.2 Design-Based Research Cycles

Design-based research refers to a systematic methodology to improve educational practices through multiple cycles of analysis, design, and implementation [116]. We followed the design-based research method to develop *UbiCoS* in three iterative cycles. This section describes each of the three cycles, including student activities and refinement of each design-based research study cycle. In the following sections, we also describe the study participants and different measurements used as a part of the help-giving analysis.

3.2.1 Cycle 1: Ratios and proportions

The first cycle served as a baseline for the work. Throughout the week, students completed activities related to ratios and proportions. For example, students worked face-to-face in teams to find the perfect ratio of red and blue paint to make purple paint and documented their work on whiteboards. They then posted photos of their whiteboards and engaged in Gallery discussions on ModelBook. Students also used the Chat feature on ModelBook to have a whole-class online conversation. Whole class face-to-face board meetings often followed these digital discussions. Students also used a tool in ModelBook called Paint Splash Phet (https://phet.colorado.edu) to model their understanding. On the fourth day, students applied their knowledge of ratios and proportions to model a moving car's speed. Using small electric cars, students measured how much distance the car covered over a set period of time. Students recorded their data on whiteboards and submitted photos to another Gallery on ModelBook. Three out of the five days, students were assigned to watch videos on Khan

Academy and participate in the online discussion as homework. The teacher revisited the Khan Academy homework during class the next day and gave students time to complete the homework if needed. Throughout the five days, the teacher emphasized "Talk Moves" that students should use to participate in the discussion to make constructive posts, such as disagreeing with another student's post and explaining why. Table 3 provides some examples of the Talk Moves used in the class:

Table 3: Example of Talk Moves.

Category	Sentence Starters
I agree and why?	I agree with your answer, because ...
I disagree and why?	That's a good point, but ...
Add on	I think you are right, but I also think ...
Ask a question	What made you think ...?
Ask a clarifying question	What do you mean by ...?

To summarize, the first cycle included two ModelBook Gallery Discussions, one ModelBook Chat, and three Khan Academy posts. The first cycle did not include activities with the teachable agent.

3.2.2 Cycle 2: Volume and surface area

This cycle aimed to implement changes based on Cycle 1 to improve learning gains, improve interaction quality, and further explore differential participation patterns across the different platforms. Based on our results from Cycle 1, we made the following changes:

1. **Minimized contextual shifts.** We designed the curriculum to minimize the number of contextual shifts (e.g., moving from face-to-face to digital discussion), helping with coordination and maximizing instructional time.
2. **Increased Khan Academy activities.** To give students more opportunities to engage on Khan Academy, we asked students to create two posts per video. We also added a Khan Academy portal within ModelBook where students could log their posts.

3. **Revised ModelBook Discussion.** Instead of having students engage in several different discussions during the Gallery Chat, we put them in a single discussion group and had them discuss a few images with a set group of students. We believed this would prompt higher-quality conversation during the digital discussion.
4. **Implemented badges.** We implemented badges as a part of positive feedback to prompt higher-quality conversation. We began awarding badges when the system detected high-quality posts in the ModelBook and Khan Academy contexts. There were a total of 8 badges facilitating students' help-giving: Explanation, Question, Relevance (domain-related posts), Feedback, Suggestion, Reflection, Co-Construction (building on someone else's idea), and Social. For example, if a student said, "How do you find the volume of a hemisphere again?" they would be given a "good question" badge. The badges acknowledged many types of contributions. For example, if a student was not sure how to answer a question but posted "good job" or "nice work" in an attempt to be encouraging and keep the conversation moving, they would receive the "Social" badge. Badges were awarded based on simple keyword matching.
5. **Introduced the Teachable Agent.** To provide students with an additional opportunity to practice and develop help-giving skills, we added tutoring a teachable agent to the curriculum.

In Cycle 2, the curriculum covered calculating volume and surface area. Following a similar pattern to Cycle 1, students engaged in a variety of online and offline activities. On the first day, students worked with the Teachable Agent to solve six problems on ratios, in an attempt to connect Cycle 2 material to Cycle 1 and give students practice with help-giving. Over the course of the week, students took measurements of cones, spheres, and cylinders and had multiple discussions (face-to-face and digital) comparing the different shapes' volumes. At the end of the week, the embodied activity asked students to calculate how many cone and spherical-shaped cups of punch could be served from a hemisphere-shaped punch bowl. Throughout the week, students were asked to watch videos and participate in discussions on Khan Academy. Through the collaboration activities, the teacher encouraged students to participate in productive discussions using guidelines based on a refined set of "Talk Moves" that incorporated our coding scheme and the badges we had designed. In this cycle, we had two ModelBook Gallery Discussions, two ModelBook Chats, four Khan Academy assignments (two posts per assignment), and one interaction with the Teachable Agent.

3.2.3 Cycle 3: Functions

This curriculum focused on functions, and we continued to examine how refinements to the activities and technology affected student behaviors. After Cycle 2, we believed we had a curriculum strategy that fostered learning, but there was room for improvement in student interaction. Based on the previous results, we made the following changes:

1. **Better distribution of classwork and homework.** Previous cycles included in-class ModelBook discussions and the Khan Academy posts as homework. In Cycle 3, we had three Khan Academy assignments (two homework, one in class) and three Gallery discussions (one homework, two in class).
2. **Improved instructions and in-class explanations for using the tools.** We looked at the problems that students had with the interface in Cycle 2 and added instructions within the interface to benefit students. We also worked with the co-designer teacher to create prompts for improved engagement during class time.
3. **Improved badge salience, automated badge assignment, and integration.** A major refinement in Cycle 3 was changing the visual characteristics of the badges and adding text that connects with the "Talk Moves" discussed with students in earlier cycles. We made the badges more visible, adding the ability to hover over the badge and see information about its meaning and how to earn it. We also improved the algorithm to increase the accuracy of badge assignments.

In Cycle 3, students continued to work in their face-to-face groups and have digital discussions using ModelBook. Students began with the embodied activity on the first day, measuring the diameter and circumference of various circles and representing their data on a graph. They used this data throughout the week in discussions about linear functions and eventually discovered that the slope of their lines was equal to pi. Students worked with the teachable agent on the same concept and prepared for this experience by reviewing the problems they would teach the agent for homework the night before. Other homework assignments included Khan Academy participation and participation in a Gallery discussion. We removed the Chat to allow more time for the Gallery Discussion and to better balance activities between classwork and homework. In Cycle 3, we had three ModelBook gallery discussions, three Khan Academy assignments (two posts per assignment), and one interaction with the Teachable Agent.

3.2.4 Method

The three cycles of the design-based research were conducted at least a month apart, following a similar procedure for all the cycles. Each cycle was designed as a five-day design study as part of regular classroom practice with students completing pre- and post-domain and motivational assessments before and after the study, respectively. The studies were designed to explore how student interactions differed between the three platforms regarding quality and quantity.

3.2.4.1 Participants

The participants of the study came from a middle school in the southwestern United States with a minority enrollment of 47%. The three cycles took place in an 8th-grade math classroom with the same group of students (N=28) and were part of the regular curriculum and done in collaboration with the teacher.

Participants Details for *RQ1-RQ3*: In the first cycle, we received parental permission for 20 students and thus excluded the other 8 from the analysis for research questions *RQ1-RQ3*. Student ages ranged from 12-14; 11 males, 8 females, and 1 student selected "Other" on the question. Of the 20 students who consented to participate in the first iteration, only 16 students participated in all elements of the study: two students incorrectly filled out the motivation questionnaire (i.e., selected multiple items), one student was absent for the pre-survey, and one student did not post on Khan Academy. Thus, the final data analysis was done with 16 students.

Participants Details for *RQ4*: The number of students who were present and consented to participate differed from one cycle to another, with 20 in Cycle 1, 26 in Cycle 2, and 24 in Cycle 3. Out of the 26 total participants, 15 were male, 9 were female, 1 "Other" and 1 did not report. Participants self-reported ethnicity was as follows: Hispanic (n=8), White (n=8), African American (n=1), Native American (n=2), Other (n=5), and Unknown (n=2).

3.2.4.2 Procedure

In addition to introducing the CSCL environments to the classroom, we also introduced Modeling Pedagogy. Based on conversations with the classroom teacher, students were not regularly exposed to interactive or technology-mediated activities in their mathematics curriculum, and her typical teaching style was more didactic. Thus, in addition to introducing new technologies, we were introducing a new culture to the classroom. Each cycle followed a similar pattern:

1. Curriculum covered five days of instruction (Monday-Friday), approximately one hour per day,
2. A domain pretest, motivation survey, and demographics survey were administered the Friday before we came to the classroom, and
3. A domain posttest and motivation survey were administered the Monday after we completed the curriculum.

During the five days of instruction, classes were taught by the co-designer teacher with the classroom teacher present and largely responsible for classroom management.

The classroom was arranged so that students sat in small groups, and each student was given their own computer. Students were assigned new small groups with each cycle. In addition to the major platforms described above, students engaged in multiple types of activities and interactions, such as: including whole-class discussion, receiving direct instruction from the facilitator, working in small groups of two or three, and individual work within ModelBook. In other words, students transitioned from face-to-face discussions to technology-mediated discussions, answering questions on Khan Academy, and practicing explaining their reasoning with the teachable agent.

3.2.4.3 Measures

Domain Assessment. The pretest and posttest consisted of two isomorphic forms designed to assess students' ability to solve proportional and ratio problems and relevant proportional definitions. Assessments were created in a collaborative process between the classroom teacher, co-designer teacher, and researchers to align the assessments with the content and state standards. *In Cycle 1,* each test form included 12 items, with eleven items assessing students' mastery of the domain concepts and one item asking students to provide an explanation. Forms were counterbalanced across participants (i.e., half the participants received form A for the pretest and B for the posttest, and half received form B for the pretest and A for the posttest). After giving the tests in the study, we noticed that a multipart question (consisting of 3 items) on Form B was unclear to students and resulted in a disproportionate number of incorrect responses. We excluded that question from the test analysis and the corresponding question on Form A. Thus, a total of 8 items were summed to assess student domain learning, with 1 item used to assess student explanatory skill. Similarly, Cycle 2 consisted of 7 questions, and Cycle 3 consisted of 8 questions.

Motivation Pre-Measure. *In Cycle 1,* we surveyed students about their attitudes towards math and mathematical self-efficacy. The instrument consisted of 22 five-level Likert-type items. Value and enjoyment of mathematics were assessed using a portion of the Attitudes Towards Math Scale [104], modified by reversing some items to balance positive and negative statements. To examine students' mathematics self-efficacy, we adapted items from the Motivated Strategies for Learning Questionnaire (MSLQ) [89, 39]. The MSLQ scale is generic, so we modified the items to be specific to mathematics. An example item is, *"I believe I will receive an excellent grade in math class."*

In Cycle 2 and Cycle 3, we asked different motivational questions. However, we did not use those in any analysis. Hence the description of those items is left out of this thesis.

Motivation Post-Measure. *In Cycle 1,* the post-intervention motivation scale consisted of 15 questions based on Expectancy-Value Theory, with 5 equivalent questions for each platform (ModelBook, Khan Academy, and face-to-face interaction). We wanted to assess whether students' perceptions of the tasks differed between platforms and varied based on their experiences during the intervention. The scale was modified from [13] to reflect students' motivation towards help-giving in math. Two example items are: *"I'm certain I can make others understand the most difficult material presented in the question"* (expectancy) and, *"I enjoy helping others with their math questions"* (value). For Cycle 2 and Cycle 3, 5 additional statements were added as the Teachable Agent was introduced in these two cycles.

3.3 Investigating Students Help-Giving Quality and Influence of Individual Differences on Student Help-Giving

This section addresses the research questions *RQ1a to RQ1d*. While *RQ1a* to *RQ1c* describes an initial analysis focusing on the investigation of students help-giving behavior in two platforms from a single cycle (Cycle 1), *RQ1d* considers all three cycles and compares students help-giving participation across the three platforms. Another distinction is while assessing the quality of help-giving in *RQ1d*, we included both high-level help and transactivity, thus giving us better insights into students' help-giving behavior across the cycles. We describe the results for each research question in the following subsections:

3.3.1 Investigation of ***RQ1a-RQ1c***:

The first research question is *RQ1a: How do individual students' help-giving interactions vary across Modelbook and Khan Academy?* As students collaborate, their motivation in help-giving can affect their participation during collaboration. Thus our second research question is: *RQ1b: How does student help-giving behavior across different platforms predict learning and motivation?* Here, we explore motivation in the context of expectancy-value theory [122]. Both expectancy (whether an individual expects to be able to perform a task) and value are essential motivational factors that might help us to understand student help-giving behaviors and design ACLS systems for multiple platforms. In addition to the influence of different learning environments, students' collaborative help-giving behaviors can also be influenced by their motivations. We considered both student attitudes towards mathematics (ATM) and self-efficacy (SE) as motivational factors, as both ATM and self-efficacy play an important role in how students learn mathematics. To support a student with a cross-platform ACLS, we need to explore how these individual motivational differences influence student interaction across different educational platforms. Our third research question is: *RQ1c: How do individual differences (motivation and prior domain knowledge) predict student help-giving behavior across multiple platforms?*

We aimed for an initial analysis using the digital interaction data obtained from the first cycle. We believe this gave us an opportunity to understand students' collaborative interactions in a middle school mathematics classroom where students have different collaborators and use different learning environments that influenced our design decisions for future cycles. In this cycle, students collaborated on two platforms: Modelbook and Khan Academy. Modelbook allows synchronous communication through different tools to promote collaboration, mainly in the form of discussion and text-based chat. On the other hand, at Khan Academy, students participate in a collaborative activity by answering questions. We investigate our research questions within the context of these two platforms.

3.3.1.1 Help-Giving Data Coding

We coded the digital interaction data using a coding scheme based on [121] with the following dimensions:

1. Level of Relevance to the content (LOR): LOR was coded using three categories:
   - *General:* information on the content but not enough to call it an explanation; e.g., "I agree because my board also was not an exact pattern."
   - *Specific:* information specific to the content; e.g., "I think the unit rate is not 2/3 but it is 2:3."
   - *Offtopic:* irrelevant to the domain content.
2. Level of Elaboration (LOE): LOE coded for on-topic (general & specific) utterances has two categories:
   - *Non-Elaborated:* answer without example or explanation; e.g., "I agree our car also did not go in a straight line."

- *Elaborated:* answer with example, proper explanation with reasoning and justification; e.g., "if we have 2 cups+3 cups that would = five but we need 20 cups".

3. Social factors (S): Finally, we classified an utterance as social if it had at least one of the following four factors:
   - *praise:* e.g., "the graph is good"
   - *apologetic:* e.g., "No offense but this makes no sense to me, sorry."
   - *polite:* e.g.,"Thank you"
   - *encouragement:* e.g., "Just do your best"

A second rater independently coded 17% of the dialogues with LOE (kappa=.805), LOR (kappa=.954), and Social (kappa=1.0). Disagreements were resolved through discussion.

3.3.1.2 Results of ***RQ1a-RQ1c***

For the analysis, we computed the total numbers of each code dimension and student-level percentages with respect to the total utterances for each dimension. Table 4 shows the means and standard deviations for N=16 for Modelbook (MB) and Khan Academy (KA):

Table 4: M and SD for each coding category.

	LOE				**LOR**					
	Non-Elaborated		Elaborated		General		Specific		Offtopic	
Platform	M	SD	M	SD	M	SD	M	SD	M	SD
MB	8.5	4.89	1.06	1.237	5.81	4.490	3.75	3.0	1.375	2.156
KA	2.438	1.364	0.81	1.223	0.25	0.683	3.0	1.155	0	0

RQ1a. How do individual students' help-giving interactions vary across Modelbook and Khan Academy?

Table 5 shows mean percentages and standard deviations of categories elaborated, specific, and social utterances for both Modelbook and Khan Academy with respect to the total utterances for each dimension (i.e., LOE, LOR, and S).

To investigate differences in interaction between platforms, a repeated measures MANOVA was conducted with percent elaborated, percent specific, and percent social as dependent variables, and platform (Modelbook or Khan Academy) as an independent variable. The

Table 5: M and SD for distinct types of utterances.

	Modelbook		**Khan Academy**	
	M	SD	M	SD
Elaborated	10.7	12.5	22.9	34.2
Specific	44.0	26.7	92.7	20.2
Social	21.9	18.2	0	0

overall model was significant, $F(3, 13) = 32.136$, $p < .001$. Univariate tests revealed that while the percent elaborated was not significantly different between conditions [$F(1, 15) = 2.480$, $p = .136$], percent specific was [$F(1, 15) = 45.226$, $p < .001$], as was percent social [$F(1, 15) = 23.122$, $p < .001$]. It should be noted that interaction at Khan Academy followed a fairly uniform pattern, with nearly all on-topic utterances being specific, and no utterances being social.

As students gave both elaborated help and specific help in Modelbook and Khan Academy, we computed correlations between elaborated help across both platforms and specific help across both platforms. Elaborated help in Modelbook was not significantly correlated with elaborated help in Khan Academy [$r(16) = 0.433$, $p = 0.094$]; and specific help in Modelbook was not significantly correlated with specific help in Khan Academy [$r(16) = 0.261$, $p = 0.328$]. Interestingly, specific help in Modelbook was correlated with elaborated help in Khan Academy [$r(16) = .746$, $p = .001$]. This analysis demonstrates that interaction was generally different across the different platforms, but for each student, interaction on one platform did not predict their interaction on another platform.

While behaviors were different across the platforms, perceptions of students' interactions in the platforms were not. A repeated measures MANOVA was conducted with each of the motivational post-measures (self-efficacy, importance, interest, utility, and cost) as the dependent variables and the platform (Modelbook or Khan Academy) as an independent variable. The overall model was not significant [$F(5, 11) = 1.082$, $p = .422$] and there were no significant univariate effects. Table 6 summarizes the result.

RQ1b. How does student help-giving interaction across different platforms

Table 6: M and SD for post-motivational measure on help-giving interaction.

	Modelbook		**Khan Academy**	
	M	SD	M	SD
Self-efficacy	3.25	1.065	2.88	1.147
Importance	3.50	1.095	3.38	1.408
Interest	3.19	.911	3.13	1.147
Utility	3.62	.957	3.38	1.025
Cost	3.75	.931	3.44	1.031

predict learning and motivation?

Table 7 shows the means and standard deviations of the pre-test and post-test scores. We conducted a repeated-measures ANOVA and found that learning was not significantly different from the pretest to the posttest. Despite the overall lack of learning gains, we still look at predictors that may contribute to learning for individuals.

Table 7: M and SD for domain assessment and pre-motivational measures (Self-Efficacy, Value, Enjoyment).

Measures	**Pretest**	**Posttest**	**Self-Efficacy**	**Value**	**Enjoyment**
M	4.563	4.687	2.742	3.312	2.617
SD	1.4127	1.4477	.8773	.7288	.975

We did a stepwise multiple regression analysis with percent elaborated in both Modelbook and Khan Academy, percent specific in both Modelbook and Khan Academy, percent social in Modelbook, pre self-efficacy, attitude towards math score, and pre-test score as predictor variables, with the post-test score as the dependent variable. The model that emerged from the stepwise analysis contained only the percent elaborated in Modelbook (β =0.584; p = 0.003) and pretest score (β =.488; p = 0.010) as significant predictors, together explaining 67% of the total variance (Adjusted R-square = 0.619; $F(2,13)$ = 13.181; p = 0.001). Thus, the only behavioral variable that predicted post-test score was the level of elaborated help in the Modelbook.

RQ1c. How do individual differences (motivation and prior domain knowledge) predict student help-giving behavior across multiple platforms?

Table 7 shows the means and standard deviations of the pre-motivational measures: math self-efficacy, value, and enjoyment. We conducted two multivariate regressions to determine how motivation and prior knowledge predict student help-giving behaviors. The first analysis was done for Modelbook behaviors. We used percent elaborated, percent specific, and social as dependent variables with pre-test scores, average self-efficacy, and average attitude towards math scores as predictors. No significant model emerged from it. Univariate tests also did not show any significant results; $F(3,\ 10) = .471$, $p = .709$, for pre-test score; $F(3,\ 10) = 1.046$, $p = .414$ for average pre self-efficacy, and $F(3,\ 10) = 1.007$, $p = .430$ for average attitude towards math score. Multivariate analysis done for Khan Academy behaviors with percent elaborated, percent specific as dependent variables with pre-test score, average self-efficacy, and average attitude towards math score as predictors also demonstrated similar results. Univariate tests didn't show any significant results; $F(2,\ 11) = .618$, $p = .557$, for pre-test score; $F(2,\ 11) = .596$, $p = .568$ for average pre self-efficacy, and $F(2,\ 11) = .286$, $p = .756$ for average attitude towards math score. Students' motivation prior to the intervention did not have an effect on their behaviors during the intervention.

3.3.1.3 Discussion

To design adaptive support for collaboration, student activity history in the collaboration contexts and current engagement in collaborative activities are essential [87]. In this section, we examined whether student interactions differed across different technological platforms, how their interactions predicted learning and motivation, and how their interactions were informed by their individual characteristics. We found that students displayed better help-giving behavior in Khan Academy compared to Modelbook, but only help-giving behaviors in Modelbook predicted student learning. Individual characteristics like prior knowledge and math motivation did not predict how students gave help.

One interesting finding from this work was that while students gave more high-quality help in Khan Academy than in Modelbook, only the elaborated help in Modelbook was predictive of student posttest scores (controlling for the pretest). The affordances of Khan Academy (asynchronous communication with an external community) may have led students to take more time to formulate their response [123], leading to more specific help and more elaborated help. In contrast, Modelbook represented synchronous, informal communication with peers, leading to overall less high-quality help but more social behaviors (which have shown in other work to be beneficial for learning [80]). In Khan Academy, because of the increased pressure of asynchronous public posts, students may have engaged in knowledge-telling behaviors [93], where they gave help on concepts they had already mastered. This may have led to less learning than their more off-the-cuff interactions in Modelbook, which may have represented knowledge-building, where they construct their knowledge as they are constructing their explanations. One implication of this finding for the design of adaptive support is that it may be sufficient to encourage more elaborated help in Modelbook to improve outcomes from help-giving. However, in Khan Academy, it may be necessary to directly scaffold students in constructing the elaborated help so that they engage in reflective knowledge-building behaviors.

Another critical element of our results is that while context dictated how students gave help, individual differences did not. Students' help-giving behavior was more elaborate and specific in Khan Academy compared to their behavior in Modelbook, and for individual students, these behaviors were not correlated with each other. This indicates that student behavior in one platform does not inform how they will behave in another platform; rather, the different platforms influence how students will help each other. Additional support for this finding is provided by the fact that neither prior knowledge, math self-efficacy, nor attitude towards math predicted how students gave help in either platform. This finding implies that a model of student help-giving on one platform is unlikely to generalize to the same student's help-giving behaviors on a different platform, and context thus needs to be part of any knowledge-tracing model of help-giving.

This study has a number of limitations. First, the sample size was small. Second, the number of interactions was greater in Modelbook compared to Khan Academy due to the design of the curriculum. To adapt to this limitation, we used student-level percentages to compute the results rather than absolute counts of student interactions. Third, students did not learn as a whole from the pretest to the post-test, possibly because the intervention time was too short or our assessment wasn't sensitive enough to detect changes in student knowledge.

Nevertheless, the present research has important implications for computer-supported collaborative learning and ACLS. Students' interactions in different platforms can be used to design individualized support that facilitates productive communication across collaborative learning environments. This goal will require a cross-platform student interaction model, domain knowledge model, and motivation model for each student. An investigation is required to understand how to make predictions about student behavior within a single platform using this cross-platform interaction model, whether and how to encourage students to participate in platforms they are less comfortable with, whether and how to encourage students to transfer their skills from one platform to a different platform, and whether and how the same student should be given different kinds of support on different platforms.

This section examined students' help-giving behavior across Modelbook and Khan Academy. The findings take a step towards establishing the need for understanding cross-platform collaborative behavior. We believe these results will ultimately enhance peer collaboration as students move between platforms of interaction.

3.3.2 Investigation of *RQ1d*:

After completing all the design-based research cycles, we focused on the following research question: *RQ1d: How do the interaction quantity and quality differ across the three platforms in the three cycles?*. We believed this would help us understand whether we observed changes in students' help-giving interaction across the cycles.

3.3.2.1 Revised Help-Giving Data Coding

To measure the quality of students' digital interaction data, we developed a coding scheme and coded the digital interaction from the first cycle. The coding scheme coded the student's utterance for Level of Relevance, Level of Elaboration, and Social (Section 3.3.1.1). However, there were a couple of limitations: first, while coding, we considered each student's utterance as the unit of analysis, but we did not consider the context of the previous utterances. Second, we did not consider transactivity at that time of coding. Third, the coding scheme was helpful in understanding whether a student's utterance was elaborated or not. However, it was not clear how the utterances contributed to facilitating the digital discussion. Hence, to effectively assess students' collaborative participation, we revised our help-giving coding scheme to address: 1) the context of student utterance, 2) whether students are building on others' responses, and 3) the extent a student utterance is facilitating the discussion. The final coding scheme is inspired by elaborated help [121] and transactive reasoning [14]. It is an ordered coding scheme that could help understand students' participation across the different platforms.

For each cycle of the design-based research, we measured the quality of interactions within ModelBook, Khan Academy, and Teachable Agent. The different levels of the coding scheme are:

1. **Minimal Participation:** Does not facilitate further conversation. Examples include off-topic comments, repeating statements from other learners, or agreeing or disagreeing with a post without further explanation.
2. **Facilitative Participation:** Has the potential to further the conversation but does not include an elaborated response. It includes comments that are related to the activity but do not contain specific content, comments that provide an answer without explanation, and social behaviors (e.g., "Thank you!").
3. **Constructive Participation:** A statement involving reasoning building on a learner's previous comment or others' comments. For example, answering a question with an explanation, correcting others with an explanation, or asking a specific clarification question.

'Level of Elaboration', and 'Level of Relevance' from the previous version were considered a part of constructive participation, and the 'Social' was considered a part of facilitative participation.

For ModelBook and Khan Academy, the unit of analysis was a student's post. See Table 8 for examples drawn from the study data. For the Teachable Agent, we coded the highest

Table 8: Examples of each of the levels of the revised help-giving coding scheme.

	Modelbook	Khan Academy
Minimal Participation	"I agree." "I don't understand."	"I have no clue why." "I don't know."
Facilitative Participation	"I like the graph and the notes you added" "It's not 2/3 it's 4/3"	"yes you are correct because he was correct" it's usually 3.14 but sometimes it's 3.1415"
Constructive Participation	*In response to a graph drawn on a whiteboard.* "i think that you did it wrong because it starte's [sic] constant for the first 3 dots then it goes at a decreasing rate backwards"	*In response to "What is a constant function?"* "a constant function is basically when its always producing the same number as the outcome for most of the time"

category achieved within a given problem (e.g., if a student had a single constructive utterance, their code was constructive participation). To compute the reliability of codes, two raters independently coded 40% of the ModelBook data, with Intra-Class Correlation Coefficient (ICCs) of .91, .86, and .91 for Minimal, Facilitative, and Constructive participation respectively. For Khan Academy, two raters coded 30% of the data, with ICCs of .79, .78, and .84.

3.3.2.2 Results of *RQ1d*

Cycle 1- Ratios and proportions: Despite discussing ratios and proportions for the week, a paired sample t-test showed no difference in pretest scores (M=6.4, SD=2.0) and posttest scores (M=6.7, SD=1.9) ($t(18)=0.88$, $p=0.39$). After consulting with the teachers, we hypothesized that this result may have occurred because the curriculum contained too many elements. In the course of a week, students participated in small group discussions, board meetings, several ModelBook activities (creating data tables and graphs, discussing vocabulary, digital brainstorming), and a hands-on data-gathering exercise. Relatedly, instructional time was likely wasted transitioning between activities.

Examining the student interactions, we see that across the two Gallery discussions, each student engaged in an average of 4.23 posts per activity. In the Chat, each student made an average of 2.58 posts. Across the three Khan Academy homework assignments, students posted a mean of 1.07 posts per assignment. Overall, it seemed that students engaged in the activities as prompted; in fact, 19/20 students participated in posting on Khan Academy, and all 20 students posted in ModelBook. This suggests that the usability of the system was sufficient to enable students to complete the assignments.

As shown in Table 9, the quality varied depending on the format of the interaction. In the Chat, 51.0% of the posts fell into the minimal participation category, whereas for Khan Academy, 72.3% of the posts represented constructive participation. This data suggests that students are capable of providing constructive feedback but don't do it consistently. It also suggests that we have an opportunity to improve the design of the ModelBook discussions to facilitate more meaningful participation. For example, in the Gallery Discussions, students could view any group's whiteboard photo and make a comment. This led students to comment on many images (increasing the number of interactions), but also led to comments that were often superficial (e.g., "I agree"), minimal (e.g., "i disagree cause I dont understand") or left some images with no comments at all. Even when comments were constructive, they were often ignored because students were likely too busy flipping between photos to participate in a back-and-forth discussion. For Khan Academy, it seemed that students were more likely to participate constructively as they were preparing a complete answer asynchronously and posting it in a public environment.

Table 9: Student help-giving interaction in Cycle 1.

	# Activities	Avg Posts/ Activity	Minimal Participation	Facilitative Participation	Constructive Participation
Gallery Discussion	2	4.23 (SD=2.53)	28.4%	58.6%	13.0%
Chat	1	2.58 (SD=2.01)	51.0%	28.6%	20.4%
Khan Academy	3	1.07 (SD=0.36)	3.08%	24.6%	72.3%

We were also interested in seeing if student participation varied between contexts. **Did students participate equally across activities, or were they more likely to engage**

with one context over another? To address this question, we categorized students as high contributors for a given context if they participated more than average and low contributors if they participated less than average. Looking at the Gallery Discussions versus Khan Academy, we see that 25% of students (n=5) were high contributors in both activities, 30% (n=6) were low contributors in both activities, and 25% (n=5) of students were high contributors to the Gallery but low contributors to Khan Academy, and 20% (n=4) of students were low contributors to the Gallery but high contributors to Khan Academy. A similar pattern is seen comparing the Gallery and Chat. These results suggest that context does matter and that by creating multiple contexts, we enable more students to practice and engage in help-giving behavior.

Cycle 2- Volume and surface area: In Cycle 2, students showed significant learning gains from pretest (M=3.8, SD=1.5) to posttest (M=5.2, SD=2.2) ($t(23)=2.51$, $p=.019$). Examining the student interactions, we see that students engaged in a mean of 2.96 posts per discussion across the two Gallery Discussions. In the two Chats, students engaged in a mean of 2.09 posts per activity. In Khan Academy, students engaged in a mean of 0.80 posts per assignment. We should note that some students posted directly on Khan Academy while others wrote posts within the Khan Academy portal in ModelBook (but did not post on Khan Academy); we count both types of posts here. Compared to Cycle 1, the amount of interaction declined in ModelBook, perhaps because we made changes to the interface for digital discussions, and there was some confusion over how to use the new interface. In addition, even though the amount of participation in Khan Academy seems similar to Cycle 1, these numbers reflect a decline since we asked students to post twice per assignment in Cycle 2 versus once per assignment in Cycle 1. In fact, there was a steady decline in participation across the four days of Khan Academy (Day 1: 16 students, Day 2: 12 students, Day 3: 10 students, Day 4: 8 students), and only 18 out of 26 students participated at all in the activity. For the Teachable Agent, students completed roughly five out of the six problems. For this activity, moving through the problems faster may mean that students understood the domain better rather than being more motivated or engaged with the activity.

With respect to interaction quality, we also see a decline in Cycle 2 compared to Cycle 1 with respect to ModelBook (See Table 10). Note that there was one Chat activity where only

Table 10: Student help-giving interaction in Cycle 2. Note: 5% of the Teachable Agent data was lost due to logging errors

	# Activities	Avg Posts/ Activity	Minimal Participation	Facilitative Participation	Constructive Participation
Gallery Discussion	2	2.96 (SD=3.19)	33.9%	64.2%	1.8%
Chat	2	2.09 (SD=1.89)	59.1%	24.0%	16.9%
Khan Academy	4	0.80 (SD=0.80)	7.23%	25.3%	67.5%
Teachable Agent	1	5.0 (SD=1.1)	21%	34%	40%

13 students participated, and there was a lot of off-topic conversation, driving the percentage of minimal participation up. Even without that activity, participation quality was quite low. We had thought badges might improve the quality of participation, and in fact, students received a mean of 2.1 badges (SD=1.2), with 22 out of 26 students receiving at least one badge. However, follow-up interviews revealed that while students liked the badges, they did not remember or understand why they received them. It is possible that confusion existed because the keyword matching used to award badges was not very accurate, and the badges were not prominently featured in the interface. The percentage of constructive participation in the Teachable Agent context appears higher than in ModelBook. It is possible that some students found it easier to participate constructively in that environment, which could be considered a safer space for interaction.

We continued to observe the same patterns with respect to student differentiation across contexts. When comparing Khan Academy to the ModelBook Gallery, 23.1% (n=6) of students interacted above average in both contexts; 46.1% (n=12) had low interactions in both, and 30.8% (n=8) showed a difference in interaction quantity between contexts (high in one and low in the other).

Cycle 3- Functions: As in Cycle 2, students showed significant learning gains from pretest (M=7.1, SD=2.8) to posttest (M=8.2, SD=3.0) (t(22)=2.45, p=.023). Across the three ModelBook Gallery activities, students posted a mean of 3.75 times per activity. For

Khan Academy, there were three activities (two posting opportunities per activity), and students averaged 1.33 posts per activity (considering both posts made directly to Khan Academy and posts made within the ModelBook portal). In the Teachable Agent, students solved roughly four of the six problems. In terms of the quality of interaction on ModelBook, we saw an improvement, with 34.4% of posts representing constructive participation. In terms of quality of interaction at Khan Academy, 73.4% of students had constructive participation (which remained fairly consistent throughout the cycles), and 28.6% of the teachable agent solutions contained constructive participation. Students received a mean of 1.96 badges (SD = 1.52).

Table 11: Student help-giving interaction in Cycle 3. Note: 1% of the Teachable Agent data was lost due to logging errors.

	# Activities	Avg Posts/ Activity	Minimal Participation	Facilitative Participation	Constructive Participation
Gallery Discussion	3	3.75 (SD=2.57)	22.2%	43.3%	34.3%
Khan Academy	3	1.33 (SD=0.97)	1.6%	25.0%	73.4%
Teachable Agent	1	4.38 (SD=1.86)	49.5%	21.0%	28.6%

Thus, while Khan Academy quality remained fairly constant, ModelBook participation quality improved, perhaps as a result of our improvements to the badging system in Cycle 3. See Table 11. Finally, we continued to observe the same patterns with respect to student differentiation across contexts. When comparing Khan Academy to the ModelBook Gallery, 20.1% (n=5) of students interacted above average in both contexts; 33.3% (n=8) had low interactions in both, and 45.9% (n=11) showed a difference in interaction quantity between contexts (high in one and low in the other), again confirming the importance of providing multiple contexts for interaction.

3.3.2.3 Discussion

Our goal was to design a system that integrates technology and curriculum to support help-giving across different contexts and to understand how that system should be implemented to best serve middle school mathematics students. In each of our three cycles, we refined the classroom activities, technology features, and thus, students' interactions in the classroom and with the technology. While our DBR methodology does not allow us to make causal claims about the relationship between the refinements and the changes we observed, the results across the multiple cycles provide insights into how to design such systems. First, in order to have sufficient instruction time to observe learning gains within each cycle, we needed to minimize the number of transitions between activities. In Cycle 1, students switched between contexts multiple times per day. This coordination time took away from instruction time and likely contributed to the lack of significant pre- to posttest learning gains in Cycle 1. When we restructured the curriculum to focus on a smaller number of contexts each day, we saw learning gains improve.

Another goal of UbiCoS is to increase the quantity and quality of interactions across contexts. Overall, the number of interactions within each context remained fairly stable across the three cycles. This remained flat even when we specifically asked students to increase their number of posts. In Cycle 1, students were asked to make one post for each Khan Academy assignment; in Cycles 2 and 3, students were asked to make two posts per assignment. This suggests that students were only willing to make a certain number of posts, regardless of the assigned amount. With respect to increasing interaction quality across cycles, we see mixed results. For the Gallery Discussion, the percentage of constructive participation posts nearly doubled in Cycle 3, whereas the quality of Khan Academy posts remained relatively stable, albeit fairly high throughout. This suggests that while badges or embedded gallery instruction may have had an influence on increasing the quality of discussion in ModelBook, giving students more time (as in the asynchronous Khan Academy environment) or other task characteristics may be more important.

Finally, one of the most consistent and promising findings from our three cycles was that incorporating multiple contexts enabled more students to become high participators in the conversation. Some students participated at a higher rate than average on the Chat but lower than average on Khan Academy, and vice versa. This suggests that by including multiple contexts, we are serving different populations. Also, the Teachable Agent provided another avenue for practicing help-giving behavior. In an interview with the classroom teacher at the completion of the study, she said that providing multiple contexts for help-giving and conversation "gives voice to those who do not talk." From her perspective, the greatest benefit of UbiCoS was that it allowed for differentiated instruction and opportunities for a broad range of students to get involved.

3.4 Conclusion

In this chapter, we have described UbiCoS which includes multiple platforms for collaboration, and described the major refinements of each cycle in terms of technology and pedagogical decisions in order to improve students' help-giving participation. We developed a coding scheme to measure the quality of students help-giving and compared students help-giving participation across the different platforms. The results indicate that the platform is a much better predictor of whether and how the students give help than their individual characteristics alone. The results also indicated that certain motivational factors, such as math self-efficacy and attitude towards math, do not influence students' help-giving interaction. Hence, this encouraged us to explore further to identify factors that might contribute to students' help-giving interaction across the platforms.

4.0 Modeling Students' Help-Giving Behavior

In this chapter, we focus on modeling students' help-giving behavior. To build an effective ACLS, it is crucial to have a system that models student collaborative behavior and suggests personalized intervention according to that. The overarching research question for this chapter is *RQ2: How can we design an explanatory model using student motivation and contextual factors together and explaining students' collaborative behavior?* While predictive models can provide valuable insights into future outcomes, explanatory models offer a deeper understanding of the reasons behind those outcomes. By combining student motivation and contextual factors in our model, we aim to uncover the complex interplay between these factors and shed light on the dynamics of collaborative behavior.

Out of the four sub-research questions for *RQ2* (see Table 1 in Introduction), we delve into the first two sub-research questions that contribute to the overarching goal of designing an explanatory model for students help-giving behavior within the context of our ACLS. The first sub-research question is *2a: Which factors motivate or inhibit students help-giving behavior on different platforms?* We conducted a semi-structured interview (n=16) to answer this question. Through this interview, we aimed to explore and identify the specific factors influencing students' willingness to help their collaborators across different learning platforms. This knowledge helped us develop targeted interventions and support strategies to enhance collaborative participation.

Moving on to sub-research question *2b: How do we model students' constructive participation using individual and platform characteristics to explain their participation?* Our focus here is on leveraging the identified factors from the previous sub-question to construct an explanatory model. This model considered both students' individual characteristics and the contextual factors associated with different learning platforms. UbiCoS consists of a synchronous digital textbook (Modelbook), an asynchronous question-answer forum (Khan Academy), and a synchronous agent (teachable agent) where students practice help-giving behavior. By using the individual factors and the affordances of these platforms, we aim to develop a model that explains students' constructive participation in collaborative activities.

This chapter will present the modeling methodology, analysis, and findings as we attempt to construct an explanatory model that captures the nuances of student collaboration.

The chapter outline follows: Section 4.1 describes the different factors influencing students' help-giving behavior from a semi-structured interview. Section 4.2 describes the modeling approach to develop an explanatory model. Section 4.3 uses the data obtained from the design-based research study and builds a logistic regression model. We describe the results in the same section. We conclude with Section 4.4.

4.1 Identifying factors Influencing Students' Help-Giving Behavior

In order to model student help-giving behavior, we first focus on the question: *2a: Which factors motivate or inhibit students' help-giving behavior in different platforms?* We believe this question will help us identify the factors influencing the students' help-giving participation in the different platforms. We used the Expectancy-Value theory (EVT) framework [122] to elicit students' help-giving experience during the various collaborative activities. While EVT is widely used to explain and predict students' learning performance, very few studies have examined how this theory can be used to extract factors that contribute to students' collaborative behavior, i.e., help-giving on different platforms.

4.1.1 Method

After the third cycle of the design-based research study (described in the previous chapter), we conducted semi-structured interviews with a subset of those middle school students to understand the motivations behind students' help-giving behaviors across platforms. We wanted an even distribution of students from both the face-to-face and the digital groups.

4.1.1.1 Participants

We conducted a semi-structured interview with 16 students (Female=7, Male=8, preferred not to say=1). The participants came from a middle school in the Southwestern

United States. Participants reported their race and ethnicity as follows: Hispanic/Latino(6), White(3), Native American (1), Mexican American (1), Jewish (1), Black/African American (1), do not know (1), missing (2).

4.1.1.2 Procedure

We used the Expectancy-Value Theory framework *(EVT)* [122] to explore what motivates student help-giving across different platforms. Interviews were centered around a 5 item expectancy-value theorem questionnaire adapted from [13]. The items asked about each platform (i.e., Modelbook, Khan Academy, and teachable agent), e.g., *"I'm certain I can help others understand difficult ideas talking on Modelbook".* To aid students in answering the questions, we showed them examples of their collaborative activities from the given platform. We asked them to describe what they did to get more detailed explanations. Interviews were conducted individually, done by two researchers, and were audio and video-recorded.

4.1.2 Analysis

Interviews were transcribed and coded by two coders following a general inductive and qualitative approach [107]. We first followed an iterative independent parallel coding process, in which two coders assigned codes to the data without first consulting or creating a codebook. In the following round, a third coder compared and merged the first set of codes to create a first-pass codebook, maintained in Microsoft Excel and ATLAS.ti. From there, two of the three coders iteratively refined the codes by first coding small percentages of the data, two randomly chosen interview transcripts, independently in ATLAS.ti then coding collaboratively to resolve discrepancies. When the codebook was finalized, the data was divided in half and coded by each coder individually. Finally, both coders reviewed each other's codes. Given the high amount of collaborative coding, we considered our codes sufficient through a process of establishing agreement, where two (or more) coders reconcile discrepancies through discussion [19].

4.1.3 Interview Results

The interviews provided insight into how students perceived each platform in terms of help-giving and what factors have influenced each platform's help-giving behaviors. We classified the codes broadly into 8 categories. Each code is described using one or more representative excerpts from the interview data, demonstrating that the students' data supported each one. A summary of the codes with quotes from the interview is given in Table 12.

Help-Giving Values: This code indicates whether students stated that they find help-giving either important or unimportant, e.g., *"It does matter to help. I do think that it's important to be able to help each other"*-(I8). While most of the students stated they found help-giving important, some students preferred expert help rather than helping others by themselves, e.g., *"I mean, giving help is good. But in math, I'd rather have a professional help someone else than me"*-(I4). Also, how students value help-giving differed depending on the platforms, too, e.g., regarding giving help to the teachable agent, I5 said *"Because it's a robot thing. I would rather teach it to someone far away and help them understand it."* This demonstrates that the platform where students are expected to give help may influence students' help-giving participation. Students who are embarrassed to talk face-to-face may value giving help in an online environment. Students' value of giving help to others will influence how they will participate in a collaborative environment.

Internal Benefits of Help-giving Students explained the personal benefits of giving help to others. The observed benefits mentioned by the students were sub-categorized into the following:

- Emotional Benefit: e.g., *"and then just giving help to another person, I mean I guess makes my day a little better because you know, spread the kindness"*-(I16)
- Enjoyment Benefit: e.g., *"because I really enjoy helping people a lot. I love the way it makes me feel because I just like helping people in general"*-(I10)
- Learning Benefit: e.g., *"(Cobi) Helped me by thinking like mutual ways of getting the answer"*-(I12)
- Social Benefit: e.g., *"it's good to help them on what they didn't know. Then, people will*

be like, "Oh, well, this person helped me with the question that I didn't know."-(I7)

However, a few negative remarks were also revealed depending on the platforms, e.g., *"Because it's a robot, ... I feel like it'd have been more enjoyable if you're talking to a peer or online"* (I13).

Help-Giving Self-Concept. Students evaluated their own ability to give help positively or negatively to explore students' perceptions about their own help-giving skills. Students who reported positively were confident about their help-giving to others, e.g., *"like most problems I could kind of figure out and then I can explain it to him in a way that he would get the answer"*-(I3). In contrast, some students exhibited negative attributes related to their help-giving skills such as *"I'm not very good at explaining things to other people, in math"*-(I4), which in turn could possibly inhibit their participation, i.e., *"Most of the time people just don't believe me, or I just get it wrong, so.."*-(I4). This suggests that students with lower help-giving self-concept are less likely to give help because they cannot effectively convey their ideas or explain the solution process to others.

Math Self-Concept. Students evaluated their own math ability positively (e.g., *"I'm very good at math and I feel like it didn't really take a lot of time for me to explain to other people about the questions they were asking"*-(I10)) or negatively. Negative evaluation may include fear of being wrong or fear of criticism, e.g., *"when I'm showing them how to do the problem I feel like I might get it wrong."*-(I15). This indicates that students with lower math self-concept may feel reluctant to attempt a solution to a problem for fear of failure.

Domain Knowledge Requirement. Students explicitly say that they require domain knowledge to give good help, e.g., *"It's important to like understand the problem to make sure that you're doing it right and getting the right answer."*-(I13). While domain knowledge is linked to online participation [21], we observe that students connected domain knowledge with collaboration as well, i.e., if a student feels they have sufficient domain knowledge, they are more comfortable helping others. This code is different from *math self-concept* because even if a student believes that they require domain knowledge to give help, their negative math self-concept might inhibit their help-giving participation.

Conscientiousness. Students mentioned participating in a given collaborative activity as a part of the instruction. It is one of the Big Five personality traits [74] and reflects the tendency to be responsible and goal-oriented. For example, students who demonstrated positive conscientiousness said *"On Modelbook, the teacher, she asked us and nobody answered, but I tried to. I tried to think about it."*-(I6).

Social Factors. These codes describe the different social factors influencing students' participation:

- Familiarity: preferential help-giving to familiar people, getting nervous working with less familiar peers, e.g., *"I mostly asked questions, and if there was someone that I knew that was on Khan Academy, I would just help them instead of help someone I didn't know."*-(I5)
- Group Dynamics: group composition effecting student help-giving during face-to-face, e.g., *"the only person in my group I most likely had a connection with was Tiger. Because she would always help me through the problems, while Sheep would just be sitting there, and talking to his friends."*-(I4).

Modes of Social Interaction. These codes describe student approaches to participation, e.g., showing empathy, seeking support, and leadership. For example, some students mentioned help-giving is important because *"..I don't want them to get confused and I want them to understand what did they good or what they need to work on"*-(I3) (showing empathy). Students who prefer familiarity felt the need to show empathy to their peers during the online discussion.

Contextual Influences. This includes platform-specific feedback or a wide variety of other factors that either suppressed or encouraged the students' participation. Examples of factors that suppressed students' participation are answers present in the discussion thread, off-topic comments in the Modelbook discussion, lack of response or stale posts in Khan Academy, etc. One example that encouraged some students to participate is if their peers asked a question and the students would reply to them. Modes of interaction (i.e., written vs. verbal) also influenced students' participation. For example, *"I just feel it's easier face-to-face because you're able to see each other's expressions, tell that they're struggling, you're able to help them out."*-(I9). The suppressing factors are among the most important factors inhibiting the students' help-giving participation across the different platforms.

Table 12: Interview results: factors influencing student help-giving interaction.

Codes	Examples
Help-Giving Values	*"it's important for me because I don't want them to get confused and I want them to understand what they did good or what they need to work on."-I3*
Internal Benefits of Help-Giving	*"Because since I'm helping them, it's giving me more practice of what I'm doing. So I'm constantly teaching myself more and more when I'm helping them, whether it's a different problem or not."-I8 (learning benefit)*
Help-Giving Self-Concept	*"..like most problems I could kind of figure out and then I can explain it to him in a way that he would get the answer."-I3*
Math Self-Concept	*"I'm very good at math and I feel like it didn't really take a lot of time for me to explain to other people about the questions they were asking"-I10*
Domain Knowledge Requirement	*"I think we need to know like, the math concept before helping someone else with it, because if you don't really know the math concept I don't think you'd be able to help anyone else."-I5*
Social Factors e.g., (familiarity, group dynamics)	*"I didn't really know any of the two girls that, that were in my group. So I wasn't really like, I wasn't as social as I would be"-I16(Familiarity)*
Modes of Social Interaction (e.g., showing empathy, seeking support from peers, and leadership)	*"you also don't want to tell them that they're automatically wrong. You just want to say what they can do better, what they could improve on"-I6 (Showing Empathy)*
Contextual Information (e.g., off-topic comments, answer already present, etc)	*"I don't think there really was much of a discussion on what we talked about, because everyone was just saying really ... some things were inappropriate, some things were just completely random-I4"*
Conscientiousness*	*"On Modelbook the teacher, she asked us and nobody answered, but I tried to. I tried to think about it"-I6*

†
* Added in the second pass of the interview coding.

Many of the factors identified are consistent with prior research in online learning, e.g., self-concept behaviors are related to student learning [23], knowledge about the domain and familiarity with other participants are motivators to contribute in online discussions [51]. However, this analysis highlighted the factors most critical in our particular setting and how they might interact with each other. Our interviews also demonstrated that help-giving behaviors are not consistent across environments and that contextual and motivational factors vary depending on the platform.

4.2 Developing a Computational Model for Constructive Participation

The research question we focus on in this section is: *RQ2b: How do we model students' constructive participation using individual and platform characteristics to explain their participation?*. We focus on constructive participation as it is our coding scheme's highest level of help-giving. A model indicating whether a student will participate constructively will help us design support to facilitate this behavior. In this section, we describe the assumptions, the process, and the findings related to this research question.

4.2.1 Factors related to Help-Giving

After identifying the factors from the interview, we decided to select a subset of such factors that directly influenced students' help-giving participation in different platforms. The two self-concept behaviors indicate students' confidence in math and help-giving, respectively, which is likely to influence their constructive participation. For constructive participation, i.e., domain-related elaborated responses, students need to be confident about their math and help-giving skills. Otherwise, their fear of judgment or fear of being wrong while explaining may inhibit the students' participation. The next factor we selected is familiarity. In the case of familiarity, students' preference to give help to people they are familiar with influences help-giving participation as well. One of the students in the interviews said *"I don't talk to people unless I don't know them face to face."*-(I8). In our collaborative environment, students give help to their peers in Modelbook. In Khan Academy, the same students give help to other geographically distributed learners they might not know. So, we believed familiarity played a key role in students' participation. Finally, the fourth and final factor we selected is conscientiousness, defined as a personality trait [74] such that it reflects the tendency to be responsible and goal-oriented. Hence, a student with low conscientiousness may not participate even if the participation environment is in their favor. So, we selected these four factors as individual characteristics: math self-concept, help-giving self-concept, familiarity, and conscientiousness.

4.2.2 Assumptions

We develop our computational model based on assumptions inspired by theory and knowledge about the students' platform. We used the four individual characteristics described above in combination with a subset of platform properties, i.e., platform synchronicity and public vs. private platform, platform task-type to describe a set of assumptions about students' help-giving participation. We defined the assumptions using constructive participation (CP) and no participation (NP) because they play a critical role during help-giving. Our assumptions are listed below:

1. *Math Self-Concept Assumption:* High math self-concept positively influences constructive participation across all platforms. Conversely, individuals with low math self-concepts tend to exhibit more constructive participation in the teachable agent platform, possibly due to the reduced fear of judgment.
2. *Help-Giving Self-Concept Assumption:* High help-giving self-concept influences constructive participation either in a synchronous or an asynchronous environment. On the other hand, low help-giving self-concept influences participation in an asynchronous environment, possibly because students can take more time to formulate their responses.
3. *Familiarity Assumption:* In a private learning environment with the student's peers, a preference for helping friends or familiar faces over unknown people may lead to increased constructive participation and reduced no-participation. However, if a student has no preference, other factors might influence their participation.
4. *Conscientiousness Assumption:* High conscientiousness encourages a lower number of no participation as they are more responsible towards their assigned tasks;

The assumptions above gave us a sense of the individual-level and platform-level factors that combine/interact to explain constructive participation and no participation in the different platforms in our collaborative learning environment. We then proceeded to assemble these individual-platform interactions into a probabilistic, explanatory model for constructive participation. Given both the goal of finding an explanatory model (i.e., not just a predictive one) and the constraints of small sample size in this study, the development of the computational model was decidedly more theory-driven and less data-driven. The main idea, however, was to treat each student participation opportunity as a noisy process in which the likelihood of constructive participation depends on a set of individual- and platform-level factors. This section first describes a model-development approach constrained both by qualitative findings and small samples. Following the explanation of the modeling approach, we describe the resulting model.

4.2.3 Modeling Approach

We proceed to develop the model in two stages. First, based on the qualitative findings, we constrain the model to include only interaction terms of a particular kind. Second, due to a shortage of data, we "fix" many of these relative magnitudes based on the expert assumptions above and allow only the weighting of the terms to "float"—to be estimated empirically. Another way to say this is the model has a fair amount of expert understanding baked into fixed parameters while retaining some parametric degrees of freedom to be determined empirically. That said, the structure of the model is quite general. With the availability of more data, it would be easy to unfix the expert-based parameters.

An overview of the model is presented pictorially in Figure 4, where the likelihood of constructive participation is theorized to be affected by three combinations of individual- and platform-level variables. We formalize this model below.

Let the individual-level variables — math self-concept (M_p), help-giving self-concept (H_p), and preference for familiarity (F_p) — be collected as Q_p,

$$Q_p = \{M_p, H_p, F_p\}$$

In principle, there could be more of these individual-level factors. Platform-level variables, such as whether interactions are synchronous/asynchronous (σ_e), public/private (π_e), or with human/agent (η_e) are collected in the set Γ_e:

$$\Gamma_e = \{\sigma_e, \pi_e, \eta_e\}$$

We have used Roman letters for individual-level variables and Greek letters for platform-level variables in addition to the subscripts p and e for added clarity.

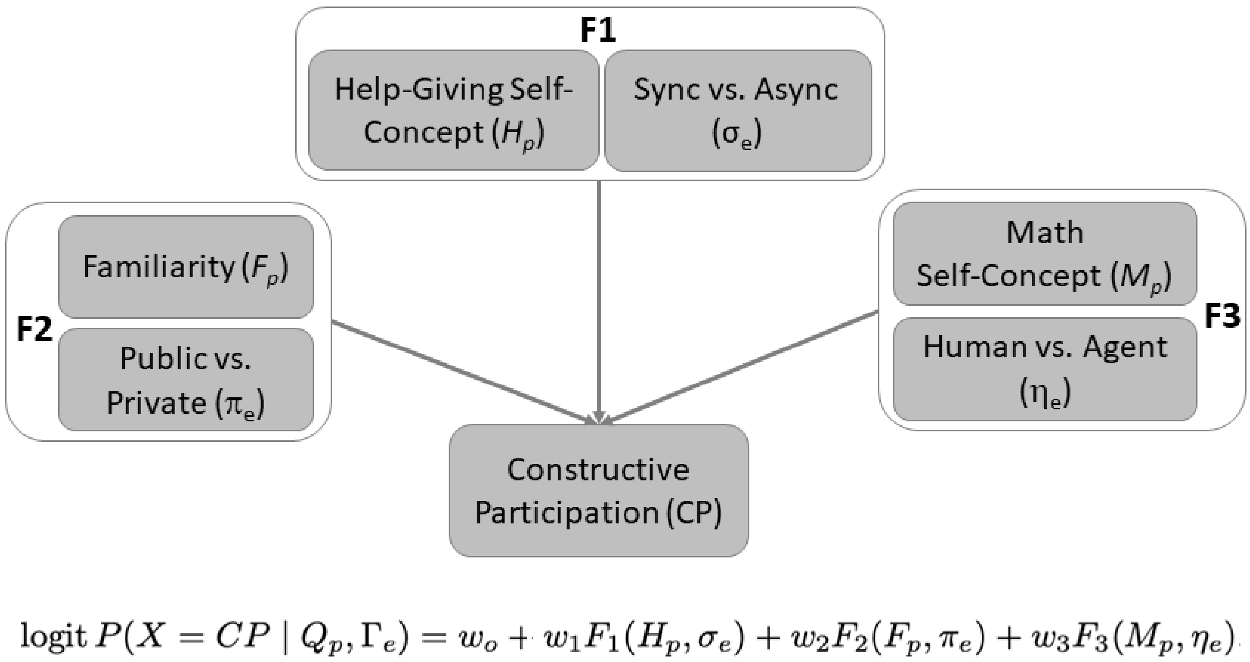


Figure 4: Constructive participation model visualization.

Finally, we combine these factors together in a probabilistic model. If we adopt a generalized linear model formalism [27], we can include the main effects of person-level variables and platform variables as well as interaction effects. With just three of each kind, this could result in six main effects and fifteen bilinear interactions. The simplifying structure we impose, which is inspired by the qualitative analysis, is to (a) formalize the model in three interaction functions and (b) assert a structure for each of these. Specifically, we write

$$\operatorname{logit} P(X = CP \mid Q_p, \Gamma_e) = w_o + w_1 F_1(H_p, \sigma_e) + w_2 F_2(F_p, \pi_e) + w_3 F_3(M_p, \eta_e). \tag{1}$$

The equation indicates that the log-odds of observing the participation variable X to be constructive ($X = CP$) are dependent on a linear combination of three functions, each of which combines exactly one person-level variable with one platform-level variable. The model expressed in Equation 1 allows for all six main effects as well as three two-variable interaction terms. It specifically excludes twelve other two-way terms (e.g., a $F_p\sigma_e$ term or F_pM_p term) and any higher order terms. The weights w_i remain undetermined. However, each of the functions can be further unpacked. Working with F_1, for example, we may write

$$F_1(H_p, \sigma_e) = a_0 + a_h H_p + a_\sigma \sigma_e + a_{h\sigma} H_p \sigma_e. \tag{2}$$

Equation 2 is a generic decomposition allowing for bilinear interactions with weights a_i. Analogous forms can be written out for F_2 and F_3. Expressed this way, however, it is not particularly natural to postulate values for the weights, even with some expert knowledge. Note that the person-level and platform-level variables here are all discrete, binary categories. The next step in development is thus to recast this form of F_1 as a two-by-two table. The values of the table are then filled in by hand. By this, we mean that we postulated simple rules (half-integer sum-rules) to quantify the four values of the table.

This process may seem arbitrary, but the precedent for it is established in the work of expert knowledge elicitation for uncertain outcomes. In this case, we treat ourselves as the experts and use our findings from the qualitative analysis to assign values.

Consider Table 13. F_1 is a function of help-giving self-concept H_p and the synchronicity σ_e of the platform. These two binary-valued variables can combine in four ways. We start out with the empty table (Table 13a) on the left. For each cell, we quantify the effect of this variable combination on the log odds of constructive participation in half-integer values. The result is the populated table (Table 13b) on the right.

Table 13: Process for expert judgment of variable interactions: the empty table is filled in with half-integer values.

F_1	σ_e = Sync	σ_e = Async
H_p = low	?	?
H_p = high	?	?

(a) Empty table before expert judgment.

F_1	σ_e = Sync	σ_e = Async
H_p = low	-1	-0.5
H_p = high	0.5	1

(b) Table with values filled in.

This process is explicitly asking for heuristic judgment. Based on the *help-giving self-concept assumption*, for example, high H_p learners are more likely to participate constructively and even more so in asynchronous environments where they have time to compose their responses. Following this reasoning, the (H_p high, σ_e Async) cell of the table is populated with 1 (a log-odds increase of 1 is equivalent to increasing the raw probability from, say, 0.5 to 0.73), while the (H_p high, σ_e Sync) cell of the table is populated with 0.5 (equivalent to an increase in probability from 0.5 to 0.62). (The negative values in the upper two cells would reduce an initial probability of 0.5 to 0.27 and 0.38, respectively from left to right). After this heuristic process is carried out, it is straightforward to convert the results back to the standard form of Equation 2. For the values in Table 13b, the decomposition is:

$$F_1(H_p, \sigma_e) = -0.5 + 1.5H_p - 0.5\sigma_e, \tag{3}$$

where H_p takes the values $\{0, 1\}$ for {low,high} and σ_e takes the values $\{0, 1\}$ for {sync, async}. Note that as a consequence of the values chosen ("fixed"), help-giving self-concept and synchronicity contribute main effects but not an interaction term to the computational model. However, this did not have to be this way, as we shall show using F_2.

To reiterate, the purpose of starting with a two-way table is that it is a more natural way for the researcher to think about interaction effects relative to one another. That is, to think in terms of individual hypothetical combinations. The functional form of Eq 4 can be "learned" (or reverse engineered) by using a regression on data generated by the two-way table (or by linear algebra). It would be awkward, even for an expert, to try to fill in the weights a_i in Eq. 2 directly.

F_2 is a function of familiarity preference F_p in {yes, no} and the public/private characteristics π_e of platforms. Our reasoning about the corresponding two-way table proceeded as follows. Based on the *familiarity assumption*, if a student prefers friends while giving help then the likelihood of participation increases in a private learning environment. However, if a student does not assert such preference, then the likelihood remains the same irrespective of the platforms. Quantities were thus assigned to the corresponding Table 14.

Note that the range of values in F_2 is [0, 2] and not [-1, 1]. A student who prefers familiarity may or may not make constructive participation in a public environment. In other words, there is no suppressing effect due to this term, so the value ranges between 0 to 2 instead of -1 to 1. We can also explain this using literature. We have seen Azmitia and

Table 14: F_2 Table with values filled in.

	π_e = Private	π_e =Public
F_p=yes	2	1.5
F_p=no	0	0

Montgomery [10] demonstrated that friends exhibit higher levels of transactive conversational moves. So, in our case, students are likely to make constructive participation in a private environment.

Converting to standard form,

$$F_2(F_p, \pi_e) = 1.5F_p + 0.5F_p\pi_e, \tag{4}$$

where, as before, the dichotomous variables assume values in $\{0, 1\}$. Here $F_p = 0$ stands for no preference for familiarity and $\pi_e = 0$ denotes a public environment. The form of F_2 does include an interaction between the person- and environment-level variables, in contrast to F_1. Note also that there is no main effect of the public/private variable. This means that, from our view of student perspective, being in a private venue was not generally an augmenting factor for constructive participation. That effect would only be experienced by students who asserted a preference for familiarity.

Finally, turning to F_3, it also takes two variables similar to F_1, and F_2. It is a function of math self-concept M_p and the different platform actors i.e., human/agent η_e. However, due to the affordances of the task within the platforms, i.e., Khan Academy requires posting a question and answer instead of a discussion thread or just a comment, we have seen students tend to make more CP in this platform. So, it took 3 values each for synchronous chat, question-answer, and communication with an agent. Based on the *math self-concept assumption*, students with high math self-concept are likely to participate irrespective of the platforms however students with low math self-concept are likely to participate more with the virtual agent. The values in Table 15 reflect that.

Table 15: F_3 Table with values filled in.

	$\eta_e = MB$	$\eta_e = KA$	$\eta_e = TA$
M_p = low	-1	0	0.5
M_p = high	0.5	1	0.5

Converting to standard form,

$$F_3(M_p, \eta_e) = -1 + 1.5M_p + \eta_{e=KA} + 1.5\eta_{e=TA} - 0.5M_p * \eta_{e=KA} - 1.5M_p * \eta_{e=TA} \qquad (5)$$

After obtaining the formal equations for each F_1, F_2, and F_3, we calculated the respective scores. We conducted a logistic regression model using F_1, F_2, F_3 as the predictor variable, and constructive participation as the outcome variable. In this model, the predictor coefficients are positive and statistically significant, indicating all the terms contribute to constructive participation. Finally, following the general equation 1 we get the following logit equation for constructive participation:

$$\text{logit}\, P(X = CP \mid Q_p, \Gamma_e) = -1.52 + 0.623 * F_1 + 0.399 * F_2 + 0.631 * F_3 \qquad (6)$$

We can use equation 6 to calculate the likelihood of constructive participation for each student at a given platform.

4.2.4 Adjustment of the Computational Model

The model in Figure 4 (Equation 1) is designed to explain constructive participation using three different pairs of individual and platform characteristics (F_1, F_2, F_3). Table 16 lists the different platform-level characteristics: we see that Modelbook and Teachable Agent almost share similar characteristics (both are synchronous and private) even though the mode of collaboration is different (text-based vs. speech-based, respectively).

Table 16: Platform Characteristics.

Modelbook	Khan Academy	Teachable Agent
Synchronous	Asynchronous	Synchronous
Private	Public	Private
Text-based	Text-based	Speech-based

So, in Figure 5, when we take a closer look into the F_1 sub-model (Help-Giving Self-Concept, Synchronicity) and the F_2 sub-model (Familiarity, Public) of the main computational model, we observe that the Synchronicity and Public are essentially the same variables, indicating that these two variables have a deterministic relation between them.

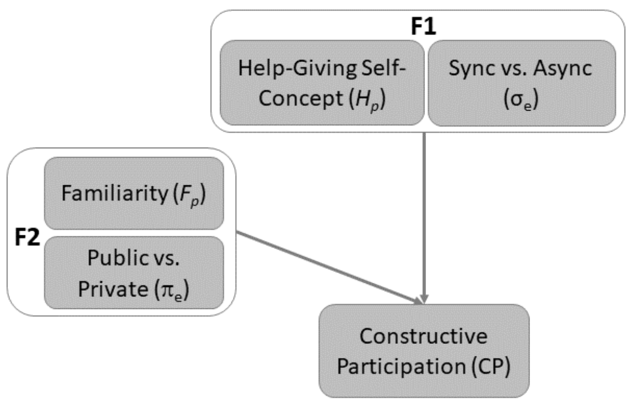


Figure 5: F_1 Sub-Model and F_2 Sub-Model visualization.

The F_1 sub-model and the F_2 sub-model are individually accurate per their definition and intended purposes. However, in the computational model presented in Equation 1, where both of these sub-models are used as predictors, the same variable is considered twice, thus influencing the interpretations of the model. To address this, we merged the F_1 sub-model and the F_2 sub-model into a new combined variable called F12_combined. This combined variable includes Help-Giving Self-Concept, Familiarity, and Synchronicity. The adjusted model is graphically represented in Figure 6. By combining these characteristics into a single variable, we aim to enhance the clarity and coherence of the model, ensuring that each predictor is distinct and contributes meaningfully to the overall analysis.

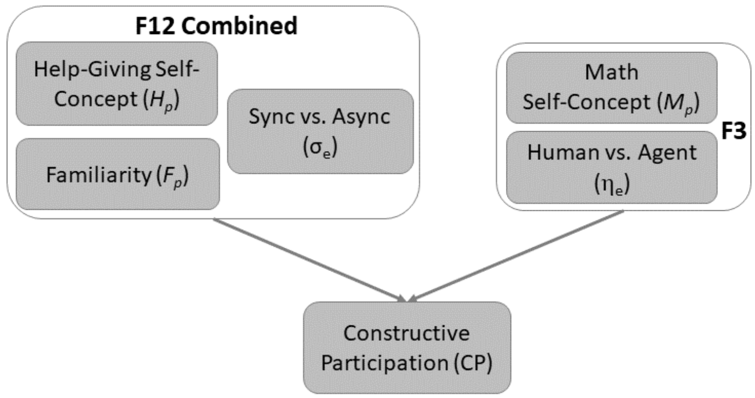


Figure 6: Computational Model for Constructive Participation.

4.3 Building and Evaluating the Computational Model

In the previous section, we used domain knowledge and expert opinion to formulate different theories and converted them into measurable variables, e.g., F_1, F_2, F_3, which are assumed to have some effect on the event of constructive participation. We described the modeling approach for developing an explanatory model using these variables. The purpose of this explanatory model is to *explain* students' constructive participation. Generally, in the case of the explanatory modeling approach, the steps involve finding a *true model* among a set of candidate models, which is needed for the purpose of explaining the data [100]. In theory, a true model refers to a parsimonious model with all the necessary variables in appropriate forms, as well as being the best model at the same time among all possible ones in terms of explaining and predicting the response variable (however, in practice, it is hard to reach a true model because of the restricted number of variables and finite sample sizes). In our case, we focus on the computational model for constructive participation and examine how well it explains the response variable. Different metrics are used to compare models using the same dataset. In this section, we briefly describe the measures that we use to compare the candidate models, then describe a general linear model (GLM) that would quantify the effect of the explanatory variables on the outcome variable.

4.3.1 Measures

In this section, we present measures related to the evaluation of the overall performance of an explanatory model. The validation of an explanatory model can be based on the evaluation of goodness-of-fit [16]. For model selection across the same dataset, we use the Bayesian Information Criterion (BIC) instead of Akaike Information Criterion (AIC) because BIC is the better criterion for finding the true model among a set of candidate models, which is what we need for the purpose of explaining the data [100]. Since the response variable is binary (constructive participation or not), we use Brier Score as a goodness-of-fit measure [15]. We evaluate the goodness-of-fit of all candidate models using Brier Score [17]. To visually inspect the model fit (how well a model matches the observed data), we use a Calibration Plot. We describe each of these measures briefly down below:

Bayesian information criterion (BIC) BIC is a method for selecting models which are under the maximum likelihood estimation framework. The BIC statistic is calculated using the formula:

$$BIC = -2 * LL + log(N) * k$$

where LL is the log-likelihood of the model on the data, $log()$ has the base-e called the natural logarithm, N is the number of observations in the dataset, and k is the number of parameters in the model, captures the complexity of a model.

Brier Score (BS) The statistic, Brier Score, proposed by Brier [17] evaluates the accuracy of probabilistic predictions (judge the quality of probability estimates). For instance, if we have two models that correctly predict the rainy weather in Pittsburgh, one with a probability of 0.53 and the other with 0.89. Assuming a 0.5 threshold, they are both correct and have the same accuracy, but the second model seems better. The Brier Score is helpful in such circumstances. The mathematical formulation of the Brier Score depends on the type of predicted variable. Since we are developing a binary prediction, the score is given by:

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2$$

where p is the prediction probability of occurrence of the event, and the term o is equal to 1 if the event occurred and 0 if not. Based on the above example, the first model predicted a 53% of rain, and it actually rained, then the model would have a Brier Score of $(0.53-1)^2$ = 0.22. On the other hand, if the second model predicted an 89% of rain, and it actually rained, then the model would have a Brier Score of would have $(0.89-1)^2 = 0.01$. The lower Brier Score indicates that the prediction from the second model was relatively accurate.

The range of the Brier Score is always between 0 and 1, with lower values indicating better model performance. The Brier Score is useful when we are interested in a model's confidence in its predictions or how probability estimates are made.

While the Brier Score evaluates the goodness of a model, the *Brier Skill Score* provides a relative metric for comparing the performance of one model with another. It is calculated as follows: Brier Skill Score = 1 - $\frac{BS}{BS_{ref}}$, where BS=Brier Score of the current model and BS_{ref}=Brier Score of a reference model. The Brier Skill Score ranges from $-\infty$ to 1, with higher scores indicating more accurate predictions than the reference model. A score of 0 indicates the same accuracy as that of the reference model.

Calibration Plot (also known as reliability diagrams in forecast literature) is a visual tool to assess the agreement between predictions and observations by plotting the predicted outcomes (x-axis) against the observed outcomes (y-axis). The idea behind a calibration plot is that when we group predictions based on their probability, we anticipate a corresponding percentage of actual events to align along a diagonal line. For instance, if we gather a set of predictions with estimated probabilities of around 10%, we would expect approximately 10% of those predictions to correspond to actual events. In the calibration plot, some points either fall above or below the diagonal line: a) below the diagonal: the model has over-forecast, the probabilities are too large, and b) above the line: the model has under-forecast, the probabilities are too small.

A calibration plot is used to investigate the prediction probabilities of a given model.

When modelers want to be confident in their predictions, they can evaluate their model via calibration plot to check that the predicted class distributions are similar to the current class distributions. Average predictions over average actuals are used to determine how close the predicted probability distribution is to the observed probability distribution from the training data. Ultimately, this shows how well the probabilistic predictions of the classifier are calibrated for binned predictions. It helps us compare two models with similar accuracy or other standard evaluation metrics.

4.3.2 Computational Model using the Design-Based Research Study Dataset

Design-Based Research (In-person) Dataset. In Chapter 3, we described the design-based research study conducted in three cycles and deployed in a middle school classroom (see Section 3.2). The study was held in-person. Hence alternatively, we refer to this dataset as in-person dataset. In this study, the students collaborated across all three digital platforms. They collaborated with their peers in Modelbook digital gallery discussions, posted questions or answered other questions in Khan Academy, and taught mathematics problems to Teachable Agent. The student utterances across these platforms were labeled as constructive participation using the help-giving coding scheme (see Section 3.3.2.1). The total number of observations in this dataset is 755 (Modelbook: 462, Khan Academy: 148, Teachable Agent: 145). There are no missing values.

Defining the GLM Model. We fit the model in Figure 6 using R's standard GLM (general linear model) function and specified family="binomial" so that R fits a logistic regression model to the dataset. We inspect goodness-of-fit with the help of the R package 'jtools' and use the 'summ' function from this package to get a summary of model diagnostics.

4.3.3 Results

In the results section, we describe the fitted model and its coefficients and describe the evaluation of the computational model.

When fitting a logistic regression model, the coefficients in the model output indicate the average change in the log odds of the response variable for a one-unit increase in the predictor variable. Often we're more interested in understanding the average change in the odds of the response variable for a one-unit increase in the predictor variable, which we can find by using the formula e^{β}. So, we convert the coefficients into odds ratio (exponentiated coefficients) and get their confidence intervals using standard error in Table 17:

Table 17: Coefficients, Odds Ratio, and Confidence Interval for the Computational Model. ** indicates p-value < 0.001, hence significant predictors.

	Coefficients	Odds Ratio	2.5% CI	97.5% CI	p-value
Intercept	-1.5405	0.2142758	0.1600037	0.2869566	0.00**
F12_Combined	0.4548	1.5758058	1.3623001	1.8227731	0.00**
F3	0.6491	1.9138923	1.4674697	2.4961223	0.00**

Results from the logistic model indicate the relationship between F12_combined and F_3 on the probability of students' participating constructively. Both of these predictors (including the intercept term) are presented to be significant predictors. It was found that holding F_3 constant, the odds of participating constructively increased by 57% (95% CI [.36, .82]) for the explanatory variable F12_combined (H_p, F_p, σ_e) It was also found that holding F12_combined constant, the odds of participating constructively increased by 91% (95% CI [.46, 1.49]) for the explanatory variable F_3 (M_p, η_e).

We compare the computational model with two other candidate models: the null model and the Brute Force Main Interaction model (Brute Force model, in short). The first reference for comparison is the null model, which is a simple, minimal model that does not include any predictor variables (intercept-only). In this model, the fitted value for each set of predictor values equals the mean of the response variable y. The corresponding slope intercept is the mean of y, and the standard deviation of the residuals is equal to the standard deviation of y. We chose the null model for comparison because it can help us to assess whether the computational model adds meaningful value beyond a basic and simple approach. If the computational model does not perform significantly better than the null model, it suggests that the additional complexity may not be warranted, and the null model may be sufficient for the given data. The second reference of comparison, the Brute Force Main Interaction model, is a model with a group of main and interaction effects between the individual characteristics (MSC, HSC, Fam) and platform type (Modelbook, Khan Academy, Teachable Agent). The number of parameters in this model is 7 (4 main effects, 3*1 interaction effects). We use this model because it uses similar variables as in the computational model but can be considered 'unrestricted' or 'untheorized' as there are no theories associated with it. Table 18 presents the BIC scores and Brier Scores for these models.

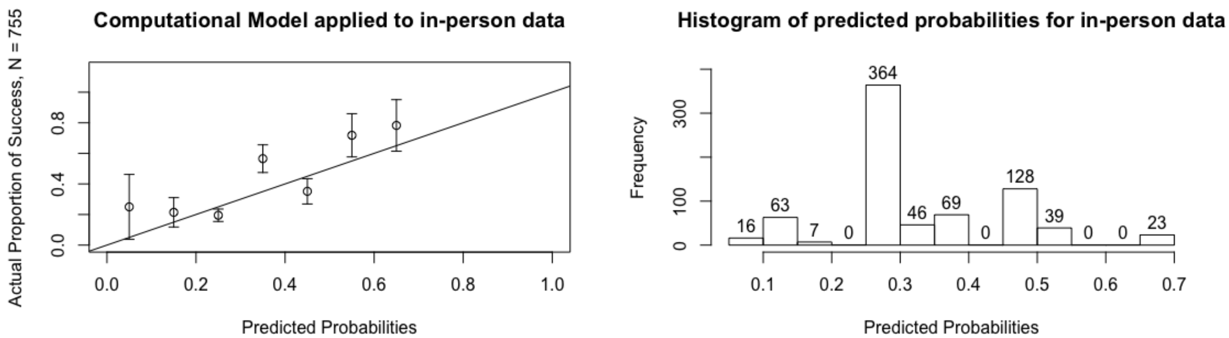
Table 18: Candidate Models including the Computational Model built with in-person dataset and their BIC, and Brier Score. The best values are indicated in bold.

Model Name	Predictors	BIC	Brier Score
Null Model	intercept-only	959.72	0.219
Brute Force Model	MSC, HSC, Fam Platform Type	959.79	0.207
Computational Model	F12_combined, F_3	**914.52**	**0.200**

The computational model with F12_combined and F_3 has a BIC score of 914.52, which is the lowest compared to the other two models, indicating a better selection of the model that describes the data. To demonstrate that the computational model is indeed better than the null model, the deviance (goodness-of-fit statistic) can be used. We examine the null hypothesis: the computational model equals the null (intercept-only) model, which is essentially a bad fit model. A p-value below the accepted decision threshold (0.05) will reject the null hypothesis indicating the computational model is not as bad as the null model. The chi-square of 58.44 with 2 degrees of freedom and an associated p-value of less than 0.001 tells us that the computational model as a whole fits significantly better than the null model.

The Brier Score of the computational model is 0.2, which indicates the accuracy of probabilistic predictions made by this model. A lower score indicates accurate predictions. To compare the models, we calculate Brier Skill Score considering the null model as the reference. We get the Brier Skill Score of 0.05 (=1-.207/.219) and 0.08 (=1-.2/.219) for the Brute Force model and the computational model, respectively. Both are positive, but the computational model has a slightly higher Brier Skill Score, indicating a better fit than the other two models.

The calibration plot of the computational model in Figure 7 helps us to inspect the fit of the model visually.



(a) Calibration Plot

(b) Histogram

Figure 7: Calibration Plot for the Computational Model and Histogram of the predicted probabilities.

In Figure 7a, after fitting the logistic regression model, each observation has an actual outcome (1 or 0) and a predicted probability (a number between 0 and 1) from the logistic regression model. We binned the observations based on their predicted probability. We have 10 bins. Any observation with a predicted probability between 0 and 0.1 go in the first bin, observations with a predicted probability in the interval 0.1 and 0.2 go in the second bin, etc. Then we calculated the actual proportion of success (1s) for observations (also known as 'event rates') in each bin. The calibration plot displayed the bin midpoints on the x-axis and the event rate on the y-axis. Each point is associated with confidence intervals.

In Figure 7b, we see the histogram plot that shows the distribution of the predicted probabilities for the in-person dataset. The histogram plot shows that a total of 677 of the observations (63+7+364+46+69+128=677 out of 755) which is 89.67% of the data, falls between the probability range of 0.1 to 0.5. Interestingly, out of these 677 observations, 364 fall into a single bin. For probabilities less than 0.1 and probabilities greater than 0.5, the model is underpredicting. Also, the bins do not cover the probabilities around 0.8 to 1 - it could be either due to the model or because there is not enough information on the data to make the predictions.

4.3.4 Discussion

The purpose of an explanatory model is to provide insights into how different independent variables contribute to the observed outcomes, allowing researchers to identify which specific elements should be adjusted to achieve the desired changes. In educational research, where sample sizes are often small, an explanatory model can play a crucial role in identifying the reasons behind observed behaviors. This holds true in the context of UbiCoS, therefore, the utility of an explanatory model can be significant. UbiCoS represents a new class of educational technology that aims to support a single student in their collaborative interactions across various platforms in which the student may have different collaborators. By developing an explanatory model, researchers can delve deeper into understanding why students exhibit certain collaborative behaviors on a given platform. This involves examining how individual characteristics interact with the features and affordances of the platform to shape student engagement. This model could provide a valuable framework for developing personalized support strategies. In recent times, developing an explanatory model for student participation is encouraged in the literature as well [95].

This chapter demonstrates the flexibility of the described modeling approach. In the beginning, we had three pairs of individual and platform characteristics. But due to an artifact of the data, we made an adjustment to the model. In terms of predicting probabilities, we observed the computational model performed well compared to the other baseline models. The logistic model developed can be used for classification using an optimal threshold; however, as a first approach to this modeling, it is better to investigate the model fit using a calibration plot and see what it tells us. For example, when built on the in-person data, the histogram in Figure 7b shows there is a frequency of 364 observations with the same probability. We have four explanatory variables for this model: MSC, HSC, Fam, and Platform Type. The Platform Type has 3 values each for the three platforms, and the rest of the variables are binary variables, so there are a total of $3*2^3= 24$ combinations. In the in-person dataset, we observe there are 18 combinations out of these 24 combinations.

One of these 18 combinations occurs 207 times, almost one-third of the data. If the model were sensitive to the bar area with these 364 observations, the model would have moved the predictions up or down. However, possibly the model is over-fitting because too much data is in this bin. If we used accuracy as our metric, we would classify a student's participation using some threshold and would have probably missed where and how the model is over-fitting/under-fitting and would have a loose opportunity to improve the model.

The modeling approach and the resulting model are interpretable, making them suitable for various studies with similar individual and platform characteristics. The model's flexibility allows for potential extension to new contexts, although further research is necessary to investigate its predictive capabilities.

4.4 Conclusion

In this chapter, we describe a semi-structured interview process where we asked the students about their help-giving experience in the different platforms of our collaborative environment. We identified a set of factors that influenced help-giving in different ways across Modelbook, Khan Academy, and Teachable Agent. While similar results are found in the existing literature, this work particularly explored how the factors specifically influence students' help-giving participation and how the factors are likely to vary depending on the platform characteristics. We then used a subset of these factors and paired them with certain platform characteristics to build an explanatory model to explain the students' constructive participation. We report an extended analysis of this explanatory model in Chapter 7 based on the data obtained from different studies described in Chapter 5 and Chapter 6.

5.0 Primary Design of Dynamic Assessment of Motivation and Adaptive Collaborative Support

In the previous chapter, we selected a set of individual characteristics derived from a semi-structured interview that potentially influenced the students' help-giving participation across the three platforms. Furthermore, we described constructing a model that incorporates both individual and relevant platform characteristics. The goal is to build an explanatory model that sheds light on the students' constructive participation across the three platforms. However, to model student participation effectively and to adequately adapt to a student's motivational state, it is important for ACLS systems to include the assessment of motivation within the learner model of intelligent tutoring systems.

A common way to assess motivation is through surveys or questionnaires. In most cases, the assessment is done at the beginning of the study, which misses two important aspects of motivation: 1) motivation is influenced by the environment [72], so it should be examined in the context of events, and 2) motivation is dynamic and changes over time [35], so it should be assessed as such. Some intelligent tutoring systems address dynamic assessment by using an interface embedded with questionnaires and sliders [30] or utilizing a single self-report question [78]. However, challenges with utilizing questionnaires to assess motivation include difficulty in understanding the vocabulary that is typically used to assess motivation [40, 44], and survey fatigue which can lower the quality of the responses/response rates [90]. Moreover, administering questionnaires during the interaction can also be intrusive [117]. Alternative non-intrusive approaches, such as using physiological data [91, 76, 98] or analyzing log files [56] are used to dynamically assess motivation. However, they can be costly, resource-intensive, and lack student agency. Student agency refers to the ability of students to make choices and take control of their learning experiences. The lack of agency can result in reduced engagement in learning activities, leading to reduced confidence in their capabilities.

The aforementioned challenges prompted us to investigate our third overarching research question in this chapter, *RQ3: How can we assess student motivation dynamically and within* *context?* The main purpose is to examine motivation within specific learning contexts, allowing researchers to understand how motivation interacts with the learning environment over time. We believe that to function in an ITS, the assessment of motivation should be **contextually embedded, dynamic, and agentic**. By integrating student agency in assessing their motivation within context, we expect to attain a more accurate representation of their motivational state. An ITS can then use this data to create personalized interventions in each context, encouraging student participation. To address the assessment of motivation within UbiCoS, we describe an application of the *Persona* method [26], a user-centered design approach for understanding important end-user characteristics like preferences and goals. It is a fictitious representative target user [26] consisting of a name, a picture, and a short narrative. We believe that our tool inspired by personas would have particular qualities that meet the aforementioned criteria, *contextually embedded, dynamic, and agentic* for assessing motivation in ITSs. Embedding the persona narrative directly into the interface may help the students to comprehend the survey items within context, making it easier to use. Furthermore, prompting the students to use the interactive tool before each digital collaborative activity enables dynamic assessment of motivation. Also, the system allows the students to modify the narrative at each opportunity as they might feel differently on different days, giving them agency in expressing their current motivational state.

We initially developed a set of personas using the four individual factors used in the previous chapters: math self-concept, help-giving self-concept, familiarity, and conscientiousness. We conducted two co-design sessions (n=13) with middle school students, where we evaluated the set of personas. We converted these paper-based personas into an interactive tool and embedded the tool in Modelbook. Subsequently, we conducted a three-week-long design study (n=17) in which the interactive tool was used by the students to report their motivation periodically. The study was conducted online due to COVID-19. The main objective of the design study was to investigate how students use the interactive persona tool. In this chapter, we focus on answering the sub-research question *RQ3a: How did the students use the interactive persona tool?* to gain valuable insights into the usability and functionality of the primary implementation of the interactive persona tool.

The other major area that we focus on in this chapter is the description of primary adaptive collaborative support. As a first step to address the research question *RQ4: How can we design adaptive collaborative support using student motivation and context?*, we designed the collaborative support for each platform using the students' individual characteristics. We implemented the primary adaptive collaborative support in the three platforms: Modelbook, Khan Academy, and teachable agent as proof of concept. However, upon reflection, the design of the support did not yield the intended results, likely influenced by certain design choices and the online nature of the study. Consequently, we decided not to analyze the usage of support in this particular chapter. Instead, our focus shifted towards reassessing and refining the support design to address the observed shortcomings, which we describe in the next chapter.

The chapter's outline follows: in Section 5.1, we introduce the Persona method and give a brief description of how this method is used in the Education domain. Section 5.2 presents the co-design sessions used to develop the personas, followed by the creation of the interactive persona tool. Section 5.3 describes the primary adaptive collaborative support using students' individual characteristics. In the next Section 5.4, we describe the design study where the students periodically used the interactive tool to report their motivational state. We then briefly present the results related to student learning and help-giving participation and report the findings of the research question *RQ3a: How do students use the persona tool?* using the data obtained from the design study. We use student trajectories to describe the changes in their motivation, which provide insight into the affordances of the persona design. We conclude the chapter in Section 5.5.

5.1 Persona in Educational Research

The *Persona* method [26] is a user-centered design approach for understanding important end-user characteristics like preferences and goals. The primary purpose of creating and using personas is to give the designers and developers a chance to get familiar with the various target users, so they can create user-centered products. Initially proposed within the context of software design [26], personas have also been used in educational research to design pedagogical interventions, e.g., [109, 97, 7, 118, 8].

Varela et al. [109] described the process of creating six child-personas (i.e., focusing on the needs of children rather than on the goal of the product [9]) for the developers and the designers to create an educational technology for children. Their goal was to design a digital math practice application tool for middle school students. Another example of the application of the personas can be seen in [97]. Sankupellay et al. [97] describe the method to develop seven personas categorizing different students of a teaching and learning support system. The personas allowed the authors to evaluate the gaps in the services of the system and make changes to fit the needs of the personas. Both Varela et al. [109] and Sankupellay et al. [97] created personas that were not directly used by the target users; however, [118] used the application of the persona method to influence student-teacher interactions and student motivation. Warin et al. [118] introduced the 'Living persona' where a teacher enacted a single persona as a part of a pedagogical approach, and the students directly interacted with the persona as a part of the pedagogy. The purpose was to improve interactions between students and teachers via the co-construction of knowledge.

In our work, we employed the persona method to create a set of representative personas based on students' motivational characteristics. These personas were then implemented as an interactive tool as an application of the persona method. The tool was embedded in the digital textbook, Modelbook, and the students (the target users) directly used the tool, modifying their motivation based on the context of their collaboration.

5.2 Primary Design of the Interactive Persona Tool

In this section, we describe the process of developing a set of personas using the students' individual characteristics in two co-design sessions. We wanted to determine how students responded to the personas as indicators of their motivation and get students' input on the persona narrative.

5.2.1 Participants

The two co-design sessions were conducted with 13 middle school students from the Southwestern United States (F=4, M=7, 2 did not report) in an after-school two-hour workshop. Participants were in 7th and 8th grade and reported their race and ethnicity as follows: Hispanic (6), Mexican (4), White (2), did not report (1).

5.2.2 Procedure

Based on the interview results described in Section 4.1.3, we initially selected four factors to develop the personas: math self-concept (e.g., confidence about one's own math competence), help-giving self-concept (e.g., confidence about one's own help-giving competence), familiarity (e.g., preference to give help to known people), and contextual factors (e.g., off-topic comments). These factors were chosen from an initial thematic analysis of the interviews with 16 middle school students about their help-giving behaviors and motivations. These factors are also related to learning in literature [23, 51]. Following the definition of the Persona method, each persona included a name, an age, a goal, a quote, and a narrative describing the persona's help-giving interactions in mathematics using these factors. Six personas (Gracie, Maurice, Sarah, Tobi, Lisa, Harry) were designed to approximate a specific type of student participation and fit the characteristics of students in our study. An example is shown in Figure 8.

In the first co-design session, each student was given the six persona documents and asked to determine how much they were or were not like the persona answering with a Likert scale ranging from 1 ("exactly like me") to 7 ("not like me at all"). 3 students rated themselves most like Gracie, 5 most like Harry, 3 most like Sarah, and the other 2 students were spread across the other three personas. This suggests that while five of our six personas resonated with at least one student, three appeared to particularly match the students in the session. Next, the students selected the persona they resembled the most and edited that persona characteristics to be more like them. Students began by modifying the persona's image, which helped ease them into the activity (e.g., making the hair longer), and continued to modify the persona's characteristics. The modifications included: (1) adding

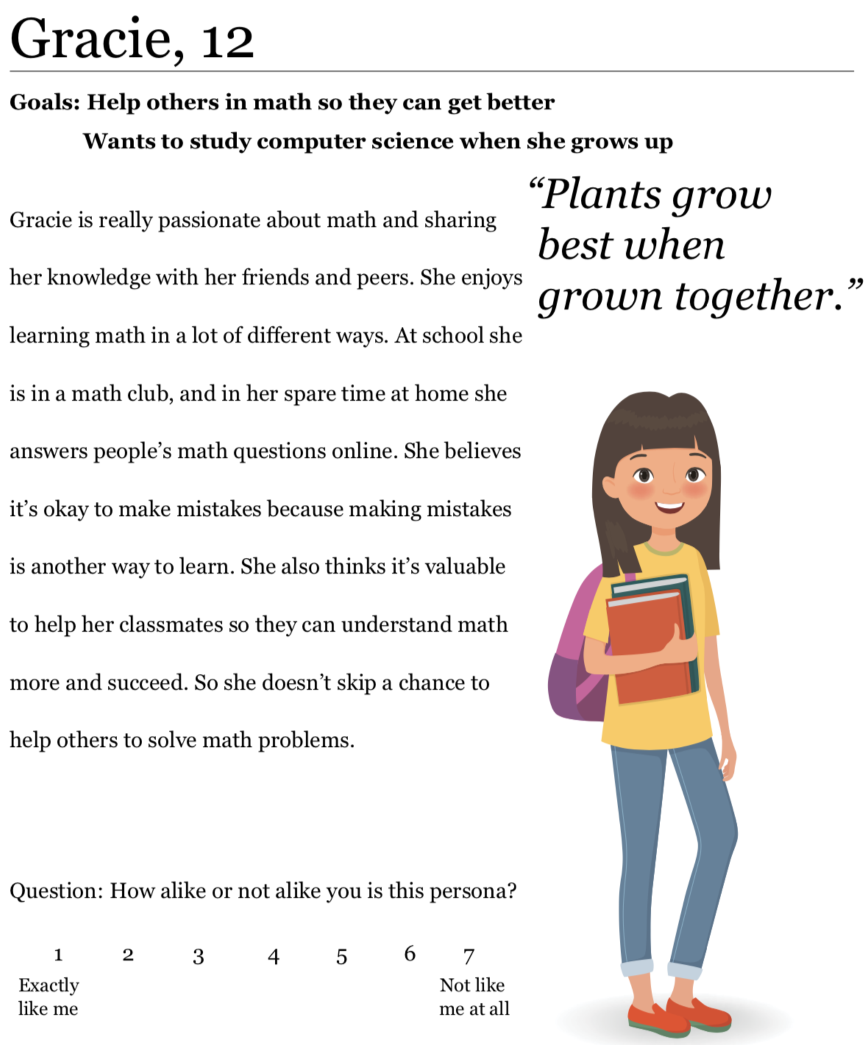


Figure 8: Persona template given to the Co-design session students.

intermediary options when talking about math performance, e.g., ‘one of the top performers’ to ‘good performer’ (7 students); (2) major editing of statements, e.g., ‘during collaboration, he fears giving the *wrong* answer’ to ‘during collaboration, he *normally gives the* answer’ (10 students); (3) minor editing of statements (5 students, e.g., modifying gender).

The second co-design session happened two weeks later with eleven students (2 from the first session were absent). Because there were many personas that students did not match to and because students made multiple edits to their persons, we decided to have students build their own personas. We gave the students a persona template with two parts: a persona narrative and a persona figure. The persona narrative included free inputs (e.g., for persona hobbies) and fixed-choice inputs with a set of options to select from related to our four factors. For example, related to math self-concept, students had three options to choose: "Really good at doing math problems", "Just ok at doing math problems", "Not great at doing math problems". The intermediary statements were inspired by the edits observed during the first co-design session. After the session, we had eleven personas created by the students and analyzed them to look for common themes, an approach often used in persona design [26]. We first used math self-concept to group the students as a determining factor in our particular learning environment, resulting in three clusters: low (4 students), medium (4 students), high (3 students). However, from co-design session 1, we observed students move from high to medium math self-concept, e.g., 'good at math' to 'almost good at math', so we combined medium and high into a single group. Then, we chose 2 personas from the low group and 2 personas from the high group such that we had at least one persona from each group with a preference towards familiarity. We chose familiarity due to its importance in designing our learning environment, which had a public and a private collaboration space. Thus, we had four representative personas, two with characteristics similar to those developed by the researchers in co-design session 1 and two more influenced by the students in this session.

5.2.3 Creating the Interactive Persona Tool

As described above, the four representative personas had a range of values of math self-concept (MSC), help-giving self-concept (HSC), and familiarity (Fam) based on student responses. We decided to eliminate the contextual factors dimension from the personas because we wanted to focus on individual motivation factors. However, we replaced that dimension with a conscientious factor based on additional analysis of the interviews. Since conscientiousness (Con) was added after the co-design sessions, we categorized each of the interviewed students under one of the four personas and then chose the level of conscientiousness that best described all the students in that persona category. The final characteristics for each of the four personas are: *Seel* (MSC:low, HSC:high, Fam:low, Con:high), *Abra* (MSC:low, HSC:high, Fam:high, Con:high), *Bellsprout* (MSC:high, HSC:high, Fam:low, Con:high), *Caterpie* (MSC:high, HSC: low, Fam:high, Con:low). Here, *Seel*, *Abra*, *Bellsprout*, and *Caterpie* are the name of the four personas.

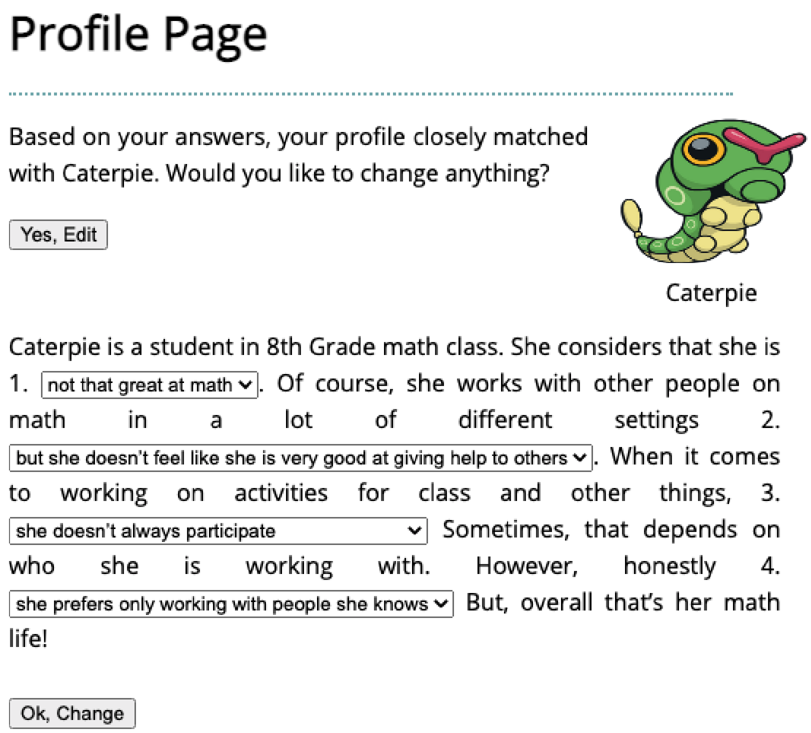


Figure 9: Primary Interactive Persona Tool interface with dropdowns for students to self-indicate their motivation.

Following the original design, we embedded the final four personas as an interactive tool in the digital textbook interface with a name, a picture, and a short narrative. The design allowed the students to modify each of the four characteristic values using a dropdown menu (Figure 9). The values are represented with words to fit in the narrative, e.g., 'pretty good at math' is mapped with high MSC, and 'not that great at math' is mapped with low MSC.

5.2.4 Discussion

The primary objective of the co-design sessions was to observe how the representative target users respond to the personas developed by the researchers. By involving the students, this approach gave the opportunity to include their perspectives and experiences, ensuring the credibility of the personas in reflecting student motivation. Through this process, we aimed to gain valuable insights into the students' motivation and refine the persona narratives accordingly.

5.3 Primary Design of the Adaptive Collaborative Support

This section describes the design and implementation of the primary collaborative support. In traditional ITS, an expert model is generally used to provide support or give feedback based on students' participation (e.g., correct/incorrect, asking for repetitive hints). Giving feedback becomes particularly challenging in ill-defined domains as solutions to problems depend on reseasoned arguments instead of a fixed number of steps. In UbiCoS, the problems students solve are open-ended, where there can be multiple solution paths. To develop a model for supporting collaboration in UbiCoS, we took inspiration from existing literature related to peer interactions [121, 45], examining several classroom techniques facilitating small-group discussion [121].

Here, we describe the peer interaction model that worked as a basis to display support to the students. The adaptive support's main goal is to improve the students' help-giving skills, i.e., giving explanations, asking clarification questions, and providing justification for their responses. Existing literature used different approaches, e.g., meta-cognitive prompts [20], question prompts [62, 124] to encourage students' collaborative participation, and these interventions were applied to all the students in the classroom. However, we want to provide personalized support that encourages participation depending on the students' characteristics. We focus on two levels for each characteristic: high and low. For example, imagine two students, one with a low math self-concept (S1) and the other with a low help-giving self-concept (S2). In order to participate constructively, S1 may need help asking clarification questions, whereas S2 may need help with explaining. A third student with high math self-concept may not need support at all to participate constructively. Inspired by the literature, we chose a specific set of peer interactions that aligns with the student's characteristics. For example, it is important for a student with a low math self-concept to learn domain knowledge to be more confident about their math skills. To solve ill-structured problems, domain-specific knowledge is required [112]. So, asking clarification questions or brainstorming the solution with others may help them learn the domain knowledge. So, for low math self-concept students, the support was particularly designed to help them ask questions or facilitate brainstorming the solution with others.

Similarly, support related to elaboration or sharing knowledge could help students with low help-giving self-concept. On the other hand, familiarity and conscientiousness are human preferences/traits. Based on the literature on self-concept and personality research, we acknowledge that self-concept behaviors are more malleable than the core personality traits such as conscientiousness and familiarity [72]. So, we want to encourage the students to participate despite their lack of interest in participating in specific platforms or completing a task for these two characteristics. We also added a fifth category called social to encourage social behaviors during the collaboration. We summarize the different peer interactions for each characteristic in Table 19. Based on the computational model described in Chapter 4, we associated the first three characteristics in the table with constructive participation and the remaining two characteristics with no participation.

Adaptive Support Implementation Description. We designed a badge for each of these peer interactions, e.g., math self-concept has three badges: Brainstorm, Question, and Critique, and each time, one of these badges was displayed randomly for math self-concept characteristics. If the students did not participate in the last two consecutive activities in a given platform, the system displayed badges related to no-participation, otherwise, it showed badges related to constructive participation. Figure 10 shows the adaptive support design from which students chose a badge before or during the interaction.

The design includes three different badges to select from, a prompt message encouraging a student to participate, and a sentence starter that is simple enough for students to edit

Table 19: Different peer interactions for four different individual characteristics.

Characteristics	Peer Interactions
Math Self-Concept	Brainstorm, Question, Critique
Help-Giving Self-Concept	Elaborate, Share, Challenge
Familiarity	Feedback, Add on, Summarize
Conscientiousness	Answer, Reflect, Assess
Social	Participate, Appreciate, Encourage

and use. In the case of constructive participation, the three badges represented interactions associated with three different characteristics MSC, HSC, and Fam, respectively. Alternatively, in the case of no participation, badges associated with two characteristics (Con, and Social) are shown. We also added 'None', allowing the students to continue participating without aiming for any badges. The system randomly selected one prompt message and one sentence starter from three alternate options so students did not see the same message. The prompt messages were related to each characteristic and varied depending on the high/low value of the characteristic. The general format for the prompt message is: [reassurance] [message] [33]. If any of the characteristics had a high value, reassurance was not added as a part of the message. The prompt messages also considered the platform differences (example in Table 20). In this design, students could copy the sentence starter into the text area using a 'Copy' button to save some typing and then complete the phrase with their own text. Some example prompt messages for each character are given in Table 20.

With this design, the students have an agency in choosing what type of post they want to make by selecting the corresponding badge. This aligns with the concept discussed by Abramovich and Schunn [2], who propose that allowing students to choose the badges they want to earn can increase motivation, particularly for learners seeking formalized recognition.

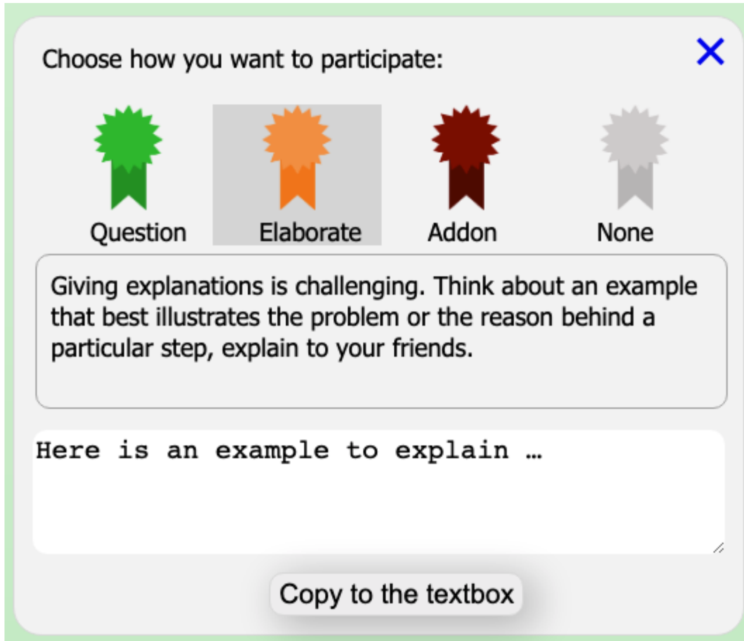


Figure 10: Primary Adaptive Support displayed at the beginning of a gallery discussion.

5.4 Design Study: Preliminary Analysis of the Interactive Persona Tool

In this section, we describe the design study where we implemented the interactive persona tool. In the results section, we present the analysis related to the research question *RQ3a: How do students use the persona tool?* and discuss the implications of such a tool in assessing student motivation within context.

5.4.1 Method

5.4.1.1 Participants

We conducted a three-week design study with 17 students (Females=7, Males=7, Prefer not to say=2, Prefer to self-describe=1). Eight participants were 12 years old and nine participants were 13 years old. Participants reported their race and ethnicity as follows: Hispanic or Latino (15), Native American (1), and Prefer to self-describe (1).

Table 20: Example of adaptive support (badge and prompt) associated with Math Self-Concept for each platform.

Characteristic	Platform	Support	Low Group Prompt	High Group Prompt
Math Self-Concept	Modelbook	Brainstorm	[Coming up with a solution is difficult!] Share your thoughts about the board so that **together** you can come up with different ideas for what to do next!	Share your thoughts about the given board so that **together** you can come up with different ideas for what to do next!
Math Self-Concept	Khan Academy	Question	[Asking questions is hard,] but if you are confused ask a clarification question **to others** to get you started on the **Khan Academy discussion**.	Asking a clarification question **to others** to get you started on the **Khan Academy discussion**
Math Self-Concept	Teachable Agent	Critique	[The more you think, the more you understand.] Based on what you learnt so far, think about alternative solutions and explain the solution to **Cobi**.	Based on what you learnt so far, think about alternative solutions and explain the solution to **Cobi**.

5.4.1.2 Measure

Motivation Pre-Measure. Math self-concept was adapted from the Programme for International Student Assessment [88], e.g., *'I have always believed that mathematics is one of my favorite subjects.'* Help-giving self-concept was adapted from Expectancy Value Theory [13] and Self Description Questionnaire-II [71], e.g., *' I'm confident that I can help others to learn the basic concepts taught in my math class.'* Familiarity was adapted from the Social Anxiety Scale for Children [64], e.g., *'I don't get nervous when I talk to new kids.'* Conscientiousness was adapted from the Big Five Questionnaire-Children [12], e.g., *'When I finish my homework, I check it many times to make sure I did it correctly.'* All the items were measured using a Likert scale ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree").

Motivation Post-Measure. The post-intervention motivation scale consisted 15 items to measure motivation to help-giving in mathematics based on Expectancy-Value Theory. The scale was modified from [13] to reflect students' motivation towards help-giving in math. Additional questions, including 5 equivalent questions for each platform (ModelBook, Khan Academy, and face-to-face interaction) were asked to capture motivation to help-giving in these three platforms. We wanted to assess whether students' perceptions of the tasks differed between platforms and varied based on their experiences during the intervention.

5.4.1.3 Procedure

In this study, all three platforms, Modelbook, Khan Academy, and Teachable Agent, were used for student collaboration. The design of these platforms was similar to the previous three-cycle design-based study described in Chapter 3. The different cycles' curriculum was restructured to fit in three consecutive weeks instead of three separate weeks. The study was designed as one hour after a class study session in Zoom. Since the study happened online, face-to-face whiteboard sessions were replaced with online whiteboard sessions. We used an online whiteboard tool, AWW (now known as Miro), and grouped 3-4 students where they solved a given problem as a group. The same teacher facilitated both the study session, as well as different small-group student discussions. At least three other researchers were present in the Zoom channel to assist with the study.

On the first day of the first week, students were introduced to the project, the research team, and the modeling curriculum. Students were asked to complete the motivational measures and demographic questions. The next day, students were given a domain assessment which was followed by a breakout room session about the 'talk moves'. In the next two days, students were introduced to the digital textbook, Modelbook, the interactive persona tool (third day), and the ratio problem. The students had a whole class discussion about the ratio problem and were sent to the breakout room as groups twice to first discuss the problem and then again to work on the whiteboards together. Once completed, students from each break-out session group were asked to upload the image to the gallery. On the fifth day, the students participated in the digital group discussion. The digital group is different than the break-out session groups. This was done so that during the digital group discussion, students from different break-out sessions can engage in discussion about the differences and similarities of their respective solutions. After the digital discussion, students were sent back to their break-out rooms to review the comments they received related to work. At the end of the first week, we have one digital discussion, one whole class discussion, and four break-out room sessions.

On the first day of the second week, students were sent to the breakout rooms to visit their groups' whiteboards, followed by a whole class discussion about the ratio problem. After that, the teacher introduced the teachable agent to the students, and the students were sent back to their break-out rooms to discuss the problem sets for the teachable agent. The students taught the teachable agent on the second day. It took the entire class time to complete this activity. The third day was a holiday, so we did not have any study sessions that day. On the fourth day, students were asked to use the interactive tool for the second time during the study. The students discussed a second problem on ratios as a whole class, a follow-up after the first problem. Students went to the break-out room to discuss the second problem. The same day students were introduced to the third and final platform, Khan Academy. The whole class viewed the Khan Academy video together and participated in the second digital discussion. The day ended with a whole class discussion summarizing the day's lecture. On the fifth day of the second week, students were redirected to Khan

Academy to make a post under the watched video. After this activity, the teacher introduced the paint splash tool to the students and had a whole class discussion using the tool. At the end of the second week, one digital discussion, three whole class discussions, one khan academy discussion, and one teachable agent interaction.

The third and the final week had a similar setup of the activities. One thing to note, based on the observation from the first two weeks, the group members were reshuffled to make a more balanced collaborative participation both in the break-out rooms and in the digital discussion. On the first day, the students discussed the third problem, i.e., collecting data across different times as a whole class. Next, students were sent out to break-out rooms to collect time vs. speed data together. The second day started with the breakout room session again to allow the students to complete the task from the previous day and also a whiteboard activity related to this problem. Students then uploaded their whiteboards into the gallery. On the third day, students participated in a digital discussion. Then the whole class watched the second Khan Academy video, and later the students were redirected to the Khan Academy to make a post. On the fourth day, students interacted with the teachable agent again. Finally, on the final day of the final week, the teacher summarized the curriculum, asked the students to use the interactive persona tool for the third and last time, and completed the post-motivational measures and the post-assessments. Students were also asked to schedule interviews if interested. The third week had one digital discussion, two whole class discussions, one khan academy discussion, and one teachable agent interaction.

In addition to the study itself, we invited students to participate in interviews to gather their insights and experiences. We conducted a semi-structured interview with a subset of the students (n=7), where we asked the students about their help-giving experiences in various collaborative activities across the platforms. We inquired about the four individual characteristics and whether they applied their help-giving behaviors. We also sought their feedback on the interactive persona tool or the badges they received during the study.

5.4.2 Results

The main objective of the design study was to investigate student interaction, and how students use the interactive tool (and the adaptive support). In this section, we briefly present the overall learning gain and student interaction data across the three platforms. We used the help-giving coding scheme described in Section 3.3.2.1 to code student utterances in Modelbook, Khan Academy, and Teachable Agent. We then present a detailed analysis of the student's usage of the interactive tool and uncover the implications it holds for understanding student motivation.

Learning Outcome. We began our analysis by first verifying that students learned between pre and posttest. For this analysis, a paired samples t-test was conducted to determine the effect of time on their math test score. The results do not indicate a significant difference between math test scores before the study (M=5.47; SD=2.748, N=15) and math test scores after the study (M=5.07; SD=3.751, N=15); [$t(14) = .658$, $p = .261$]. This suggests that there might be a lack of learning taking place. One possible reason for this could be that the limited duration of the instruction time and the presence of various concurrent activities might have posed challenges for the students in the online setup. Additionally, the learning assessment may not have been aligned with the specific type of activities in which the students engaged. During the study, the students explicitly participated in help-giving activities in various formats, however, the assessments focused on giving a correct answer to the problems or responding to multiple choice questions. This misalignment between the assessment tasks and the actual activities could have affected the learning scores.

Participation Results. We analyzed student utterances across the three platforms. After the study, we had a total of three Modelbook gallery discussions, two Khan Academy discussions, and two Teachable Agent interactions. In terms of the quality of interaction on ModelBook, we saw across the three Modelbook Gallery discussions that 45.71% of the posts represented constructive participation. Regarding the quality of interaction at Khan Academy, 72.73% of students had constructive participation, and 21.53% of the teachable agent solutions contained constructive participation.

We observed that Khan Academy exhibited the highest level of constructive participation

Table 21: Student help-giving interaction in Online Study. Note: 25% of the Teachable Agent problems were attempted but not solved, and 20.80% of the Teachable Agent problems were not attempted.

	# Activities	Minimal Participation	Facilitative Participation	Constructive Participation
Gallery Discussion	3	18.29%	36.00%	45.71%
Khan Academy	2	9.09%	18.18%	72.73%
Teachable Agent	2	4.17%	28.47%	21.53%

compared to the other two platforms. It could be due to the platform affordances where students either post a question or respond to other students' questions which inherently defines constructive participation. This finding is consistent with the previous study (Cycle 3 of Design-based-research). In the case of Modelbook, we noted nearly half of the interactions fell into the constructive participation category, which is an improvement over the previous study as well. It is important to note that we acknowledge this study involved different students. The implementation of the adaptive badge system may have contributed to increased constructive participation. Interestingly, we see a notable change in participation when examining the interactions with the Teachable Agent. Students did not attempt to solve the problems. This could potentially be attributed to the online nature of the study, where students might have encountered difficulties in using the teachable agent or faced environmental challenges such as inadequate internet connectivity or lack of access to a computer. For example, one student mentioned using the phone on the day the Teachable Agent activity happened which might have limited their participation.

The participation findings demonstrate the variation in how students provide help across different platforms, which aligns with our previous studies. These can be attributed to the different individual characteristics and their interaction with the contexts. To further explore the dynamics of student motivation throughout the study, we asked the students to use the interactive tool to report their motivation multiple times. In the remaining part of the result section, we delve into the students' usage of this tool and analyze its implications for understanding their motivation patterns.

Interactive Persona Tool Usage Results. On the first day of the study, once the responses to the questionnaire were collected, one researcher calculated the average for each factor. This was compared against a fixed cut-off value of 3.5 to mark high vs. low for each factor for each student. The following are the total number of high and low students for each factor: MSC (7, 10), HSC (13, 4), FAM (7, 10), and Con (14, 3). We calculated the number of matching factors between the students (e.g., MSC:high, HSC:low, FAM:high, CON:low) and each persona. As we had four personas, students were assigned the persona with the minimum dissimilarity of factors. If a student matched equally with two personas, the student was assigned randomly to one.

The interface allowed students to navigate to a screen to see the matched persona (Figure 11a). A student could modify any of the four values using the dropdown menu separately. If the student modified a factor and hit 'Ok, Change', the system re-calculated and re-matched with the closest persona again. Over the three-week period, the students were given specific verbal instructions by the instructor during the class (three opportunities: third, ninth, and fifteenth day of the study) to visit the interactive persona tool and make changes. However, because the instructor did not allocate sufficient time for the third opportunity, only three students viewed their personas as part of this opportunity; thus, we only report results from the first two opportunities as part of this exploration.

We answer *RQ3a: How do students use the persona tool?* by a) examining the assignment of personas to students, b) how students modified their assigned personas, and c) finally, why some students modified their personas multiple times. Because of the small sample size, our results are descriptive and intended to inform the iterative design of the tool.

How does student self-report compare to the personas? We refer to the persona matched with the student's survey response as the *initial persona.* 100% match occurs when all four factors match with the survey responses. Out of the 17 students, 10 students had a 100% match. Among the remaining 7 students (less than 100% match), 3 indicated low, and 3 indicated high for all factors, respectively. This indicates that while our personas matched the majority of the students, it may have been helpful to have two additional personas that represented the extremes on the motivational scale.

How do students modify their initial personas? To assess students' motivation, students were asked to review the persona description and, if they wanted, to modify characteristics of their *initial persona* to be more like them. If the persona and the survey questionnaire were assessing the same construct, we would expect a) students that were matched 100% to not change their persona, and b) students with less than 100% to make modifications that move the persona to be closer to the questionnaire. 9 out of the 17 students did not modify their *initial persona* even though 4 of these students had a less than 100% match with their survey responses. Among the remaining 8 students who modified their *initial persona*, 5 had a 100% match, 2 had a 75% match, and 1 had a 50% match between the persona and their survey response. To summarize, nearly 50% of the students modified their assigned personas, which indicates students may feel differently about their motivation when viewing descriptions in context compared to out-of-context survey responses.

Why do some students modify their personas multiple times? Once students modify their *initial persona* in the first opportunity, they are matched with the same or another persona (referred to as *Persona-1*) which could also have a 100% match or less based on the modification. Four students (P5, P9, P11, and P12) modified their *initial personas* as well as their *persona-1* in the second opportunity (modification in persona-1 is referred to as *Persona-2*). The average number of modifications done in the *initial persona* and the *Persona-1* amongst these students is 2.5 and 3.25, respectively, indicating their motivation may have changed over time. We observe students' trajectories across the two opportunities: in the first opportunity, students modified a factor in one direction, but in the next opportunity, the same student modified that factor in the opposite direction, e.g., P9 modified Fam from low to high first, but in the second opportunity changed it back to low. It could be because as P9 spent time within the context, his preferences for giving help changed. Second, students modified the same factor in the same direction in both opportunities, e.g., P5 modified HSC from high to low both times, which could indicate that the survey did not capture the students' actual motivational state and placing the assessment in context increased consistency.

5.4.3 Discussion

In this section, we described the deployment of the interactive tool in a design study and investigated how students used the tool to report their motivation. We embedded the personas in the interface, allowing students to report their motivation in context. This contextually embedded, easy-to-understand narrative may lead the students to respond differently than to surveys. It represents a multidimensional perspective on motivation as it suggests motivation cannot be adequately explained in terms of a single construct [72]. During the interviews, the majority of the participants acknowledged that the personas either fully or partially matched their motivation. In cases where there was a partial match, they utilized the tool to modify their motivation accordingly. This indicates that the tool effectively captured the dynamic nature of student motivation to a certain extent. By examining how the students interacted with the tool and utilized its features, we aim to evaluate its effectiveness in facilitating student engagement and understanding its potential impact on their learning experiences.

We summarize our key observations based on the results above, which later informed several design decisions for our subsequent classroom study (described in the next chapter). First, with only four personas, it can be challenging for students to see the impacts after they modify, i.e., a student might change one factor but still match with the same persona. As such, a full factorial of personas (4 factors with 2 options for each, 4^2=16 personas) could be more meaningful, and all students can have a one-to-one mapping with a persona. Second, adding a "no change" button in the tool could help understand student interaction with the tool. Without such action, it was unclear whether the students did not make changes because they were happy with their match, didn't know what to do, or got bored/distracted. This is especially important when we do online studies where it is difficult to observe students' behavior and interpret their interactions. Third, students who modified sometimes moved away from their original survey responses, which indicated the personas might be capturing dynamic changes in the students' motivation. It could also indicate the persona captured either the same or different constructs as the survey items.

5.5 Conclusion

In this chapter, we began with the co-designs to create personas for assessing motivation dynamically and in context. The students validated the factors used to develop the personas and brought their own perspectives in the process [7]. We embedded these personas in the interface, allowing students to report their motivation in context. The results indicated the personas might be capturing dynamic changes in the students' motivation. This chapter also described the adaptive support design using the students' individual characteristics. However, upon reflection, we realized the design of the support did not yield the intended results. As a result, the analysis of support usage was excluded from this chapter. It is definitely a shortcoming, but, we shifted our focus towards reassessing and refining the support design to address the observed shortcomings and enhance its effectiveness in the next chapter. In addition, due to COVID-19, we shifted the study setup to an entirely online setup. We believe this gave us insight into the student's help-giving participation in an online setup as compared to the help-giving participation in a formal classroom environment.

6.0 Revised Design of Interactive Persona Tool and Adaptive Collaborative Support

In Chapter 5, we described the development of the interactive persona tool and a three-week-long design study conducted online where the students actively engaged in collaborative activities and used the tool to report their motivation multiple times. By capturing the changes in student motivation, we gained insights into the dynamic nature of student motivation within the context of the learning environments. Additionally, the students also received personalized support based on their individual characteristics (Section 5.3).

We used the observations from the design study to identify specific aspects of the interactive persona tool and the adaptive collaborative support that required refinement and iteration, which is the focus of the current chapter. We conducted a classroom study where we implemented the revised interactive persona tool and the revised adaptive collaborative support. Similar to the design study, the students used the interactive persona tool during the study period. Hence we investigate the same research question as in Chapter 5, *RQ3a) How do students use the (revised) persona tool?* By examining the student interactions with the persona tool, we sought to determine if the findings from the previous implementations were consistent. This would emphasize the need for such a tool to assess and address student motivation within context.

Furthermore, by analyzing the data collected from the interactive persona tool, we aim to understand whether the modified characteristics better predict students' collaborative participation compared to their survey responses. We investigated *RQ3b) How does interaction with the persona tool relate to students' help-giving participation across the different platforms?* This would further support the idea that the interactive persona tool assesses student motivation and is a useful predictor of their collaborative patterns. Thus, one of the main contributions of this chapter is to demonstrate that a persona-based approach may be an effective alternative for assessing motivation within ITSs.

The ultimate goal of our work is to support student collaboration across different platforms. We continue to investigate the fourth overarching research question, *RQ4: How can* *we design adaptive collaborative support using student motivation and context?* Previously, we used only individual characteristics to provide adaptive support (described in Section 5.3). As we recognized that motivation depends on platform affordances, we revised the adaptive support system using both individual characteristics and platforms. We designed the adaptive support using a combination of prompts in the form of sentence starters and badges as a reward to encourage constructive participation.

To assess the effectiveness of the adaptive support, we investigated the two sub-research questions related to the overarching research question *RQ4*. The first sub-research question is *RQ4a: How did the students use the support across the three platforms?*. This focused on understanding how the students used the support, including the sentence starters and badges, across the three platforms. The second sub-research question is *RQ4b: What is the relationship between student support usage and student participation?*. This aimed to explore the relationship between student support usage and their participation levels. By examining this, we aimed to gain insights into how the support influenced participation behavior in different contexts.

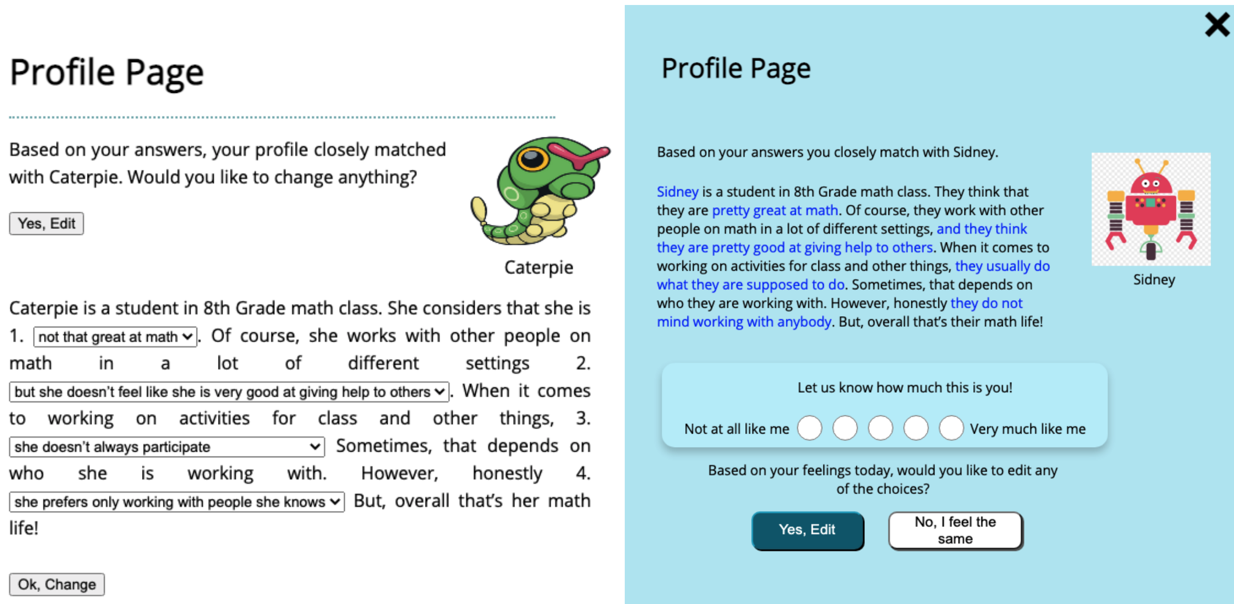
The overview of this chapter follows: In Section 6.1 and in Section 6.2, we describe the revised interactive persona tool and the revised adaptive collaborative support, respectively. In Section 6.3, we describe a classroom study where we deployed the UbiCoS system, including the tool and the adaptive support system. In the results section 6.4, we have two sub-sections: one reporting the results related to students' interactive persona tool usage following the overarching research question, *RQ3* and the other reporting the results related to students' support usage following the overarching research question, *RQ4*.

6.1 Revised Interactive Persona Tool

In the previous chapter, we examined the usage of the interactive persona tool (Figure 11a) and how students interacted with it to modify their individual characteristics during the design study. However, our analysis revealed certain challenges with the tool, prompting us to make revisions before conducting the classroom study (described later in this chapter). For instance, the limited number of personas (four) made it challenging for students to perceive the impacts of their modifications, as they could still be matched with the same persona even after making changes. Additionally, it was difficult to interpret why students chose not to make any modifications. Hence, we revised the interactive persona tool. Furthermore, the number of times the students were asked to use the tool was insufficient to thoroughly investigate the dynamic nature of motivation.

Revision of the Interactive Persona Tool In order to address the challenges identified in the previous study, we made several updates to the interactive persona tool (updated design in Figure 11b). First, we introduced 16 personas (four individual characteristics and two levels each) that allowed one-to-one mapping of factors based on the students' survey responses. Second, we added a button *'No, I feel the same'* that students can use to explicitly indicate they did not want to change any factors (i.e., remain with the same persona the system matched). The inclusion of such a button in the tool would help determine whether students refrained from making changes due to satisfaction, confusion, or disengagement. This is particularly important in online studies where direct observation of student behavior is difficult. We also added a radio button that students can use to report how much they resonate with the system-assigned persona on a scale of 1 (less likely) to 5 (most likely). Both of these changes within the interface were expected to help better interpret student interaction with the tool. Third, we used gender-neutral robot images and names instead of the Pokemon names and images used earlier. We chose gender-neutral names with an expectation to not bias the students.

Revision related to the Instruction of using the Tool We also modified the instructions regarding the use of the interactive persona tool. First, the students were given six explicit opportunities (one at the beginning, the rest of the five others before each digital activity) to visit the interactive persona tool and make changes. This ensured students had enough opportunity to modify the assigned persona and thus enabled us to investigate the dynamic assessment of motivation in different contextual settings. Additionally, with this design, we specifically asked the students to modify their motivation in relation to the context rather than in the middle of the tasks, unlike the approaches in some literature [32, 78]. We recognized that the students' motivation may remain stable within our inter-



(a) Initial Persona Tool Interface

(b) Updated Persona Tool Interface

Figure 11: Interactive Persona Tool Interface.

vention timeframe, but it may vary based on the platform they are collaborating in. By focusing on the contextual connection, we aimed to capture the dynamic nature of student motivation and its influence on collaborative activities across different platforms. Second, in addition to the verbal instruction by the teacher, the instructions were also embedded in the Modelbook. An example instruction looks like *'Today you will have a digital discussion with your classmates. As always, before you start, are there any changes you would like to make to your profile page based on how you are feeling today?'* This way, the instruction to report student motivation via the tool was made context-specific.

6.2 Revised Adaptive Collaborative Support

This section describes the revised adaptive collaborative support over the primary adaptive support (described in Chapter 5). The design of the primary adaptive support was based on the students' individual characteristics. We observed a few design limitations. First, we designed fifteen badges (three badges supporting each of the four individual characteristics and three more badges for encouraging social behaviors) that students could get during the collaborative activities. The number of badges could have been too overwhelming for the students to follow. Second, as a part of the design, the students were given the option to choose a badge they want to receive during their collaboration. However, the option to select a badge during the collaborative activity could have been disruptive to their participation. Third, the badges were designed following certain peer interaction theory [121, 45]. However, the display of the prompt to select the badges before collaborative activities lacked a clear theoretical motivation. Finally, the support only considered the individual characteristics and did not consider the platform characteristics. In order to make the support adaptive to the platforms, we revised the support design, including both individual characteristics and platforms. In the revised design, we attempt to address these issues.

6.2.1 Collaborative Support Theory

To design effective support that facilitates help-giving, we drew inspiration from the help-giving literature, where the teachers adopt different strategies to target specific student behavior in a classroom. In several studies, the researchers train the students for effective collaboration prior to collaboration. For example, Webb and Farivar [120] provided instruction on explaining three different skills: *helping skills* (e.g., asking for and giving elaborated explanations instead of only the answer, asking clear and precise questions), *basic communication skills* (e.g., checking for understanding, sharing ideas and information, and checking for agreement), and *norms for group behavior* (e.g., attentive listening, equal participation by everyone).

Building on this, we have designed three supports mapping each of the terms above:

1. **Help-Giving Support:** It has two sub-categories:
   a. Elaboration Support (ES): providing support to give elaborated responses with justifications and reasoning;
   b. Question Support (QS): providing support to ask domain related questions that will facilitate the discussion.
2. **Transactive Support (TS):** providing support to encourage students to comment on their collaborators' response, i.e., answer questions, add further information, or correct their statements if applicable;
3. **Participation Support (PS):** providing support to encourage students to participate in general.

We expect that these supports will benefit the students in developing their domain knowledge and collaborative skills. For example, elaboration support might encourage the students to give high-level help during collaboration. Giving explanations involve cognitive restructuring, which helps to understand one's own perspectives [120]. The question support could facilitate students' understanding of domain knowledge by activating prior knowledge and asking specific questions. On the other hand, transactive support is designed to allow the interplay of giving and receiving help during collaboration. Receiving explanations benefit students when the answers are elaborated and are actively used to solve problems [120].

6.2.2 Collaborative Support Description

As a part of the collaborative support, we planned to provide support in two ways across the three platforms:

1. Prompt the students to participate by displaying support before or during collaboration
2. Reward badges based on students' participation designed as positive feedback

We describe these approaches in the following two sub-sections:

6.2.2.1 Selection of Support to Display Prior/During Collaboration

We begin by providing a description of the collaborative support that is displayed to the students prior to or during their collaboration. The selection of a specific support to display to the students depends on the likelihood of the constructive participation model described in Chapter 4, the platform on which the student is collaborating, and the value of certain individual characteristics in sequence. These steps for specific support selection are given in Algorithm 1. We describe this in a bottom-up approach, i.e., the selection of support based on the high/low value of individual characteristics, then the selection of support based on the platform, and finally, the overall model for choosing the support to display prior to/during collaboration.

Motivation Based Collaborative Support The purpose of displaying support was to facilitate constructive participation. When giving support to the students, we provided support based on each characteristic's value, i.e., high and low. The variation of the support for the high and low groups of students is theoretically inspired. Figure 12 summarizes the support for each characteristic's low and high values.

Low Group Support Design: We designed support for the students in the low group to help them 1) make an elaborate response, 2) ask questions, and 3) participate in general. As students with *low math self-concept* doubt their abilities, they are likely to think asking a question publicly demonstrates they cannot solve a problem. This fear of judgment could result in avoidance of asking questions [121]. In such a case, question support could help these students to ask questions during the collaborative activity. For example, *"You can earn a badge by asking questions to express your confusion. Copy the phrase, and fill in the rest in the textbox. "I am confused about [...]. Can you explain this with an example?"* The support may encourage the students to ask a question and get a question badge contributing to the overall collaboration. Similarly, students with *low help-giving self-concept* lack the capability of providing comprehensible explanations, which leads to poorly labeled explanations or phrase explanations in confusing ways [121]. The elaboration support may particularly help these students give an elaborated response. For example, *"You can earn a badge by explaining your answer with an example about the topic. Copy the phrase, and fill in the*

Algorithm 1 Support Selection Algorithm

```
Require: likelihood ≥ 0
  if likelihood ≥ 0.3 then
      support ← Transactive
  else
      if CON == high then
          support ← Transactive
      else
          if platform == MB then
              if MSC == high then
                  support ← Transactive
              else
                  support ← Question
              end if
          else if platform == KA then
              if Fam == high then
                  support ← Participation
              else
                  support ← Transactive
              end if
          else if platform == TA then
              if HSC == high then
                  support ← Transactive
              else
                  support ← Elaboration
              end if
          end if
      end if
  end if
```

rest in the textbox. The answer is ..., since ....” The students with low conscientiousness or students who have a preference to give help to familiar faces may not feel responsible enough to participate in the assigned tasks [33] at a given platform. Hence in such cases, the participation support may encourage students to participate. For example, *“You can earn a badge by appreciating others’ effort to solve the problem.” Copy the phrase, and fill in the rest in the textbox. Good job on adding/explaining ...”* The participation support is also expected to encourage social behaviors among the group members.

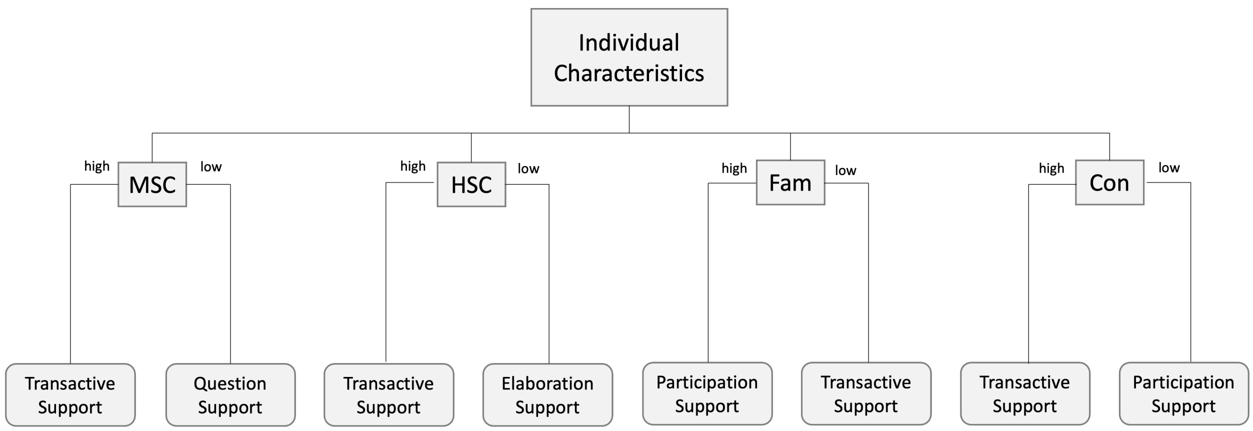


Figure 12: Support flowchart for low group and high group students for each characteristic (MSC=Math Self-Concept, HSC=Help-giving Self-Concept, Fam=Familiarity, and Con=Conscientiousness)

High Group Support Design: In contrast, for the students who are in the high group, we designed the support to encourage the students to 1) make elaborated responses or thought-provoking questions and 2) encourage others to participate. The transactive support was particularly designed for this. For example, *“You can earn a badge by expressing whether you agree or disagree with others solution. Copy the phrase, and fill in the rest in the textbox. I agree/disagree because, ... ”* This may facilitate transactive behaviors within the discussion, create an opportunity for students in the low group to practice help-giving in the form of elaboration, and also let the students in the high group act as an implicit moderator role. The benefit of a role moderator within a peer interaction is well established in literature [68].

Apart from the facilitating teacher, this approach is likely to be another source to remind the students about help-giving within the group.

Platform-Based Collaborative Support The next level of collaborative support is based on the platform the student is collaborating in. Since there are four individual characteristics, and we have platforms with different characteristics, we prioritize specific individual characteristics depending on the platforms and select the appropriate support to display to the students. The priority of the characteristics for each platform is based on the assumptions described in Section 4.2.2. The selection of the characteristics for each platform is described in the following:

Modelbook: According to the *Math Self-Concept Assumption*, students with low math self-concept are less likely to participate in Modelbook due to the fear of judgment or fear of being wrong. Hence, providing support for low math self-concept students in Modelbook is essential for effective collaboration. So, low math self-concept is prioritized in the case of Modelbook.

Khan Academy: According to *Familiarity Assumption*, students with high familiarity are less likely to participate in Khan Academy as these students prefer to help their friends or peers. Hence, participation support encouraging students to participate in Khan Academy can be helpful. So, preference for familiarity is prioritized in the case of Khan Academy.

Teachable Agent: In the case of Teachable Agent, we consider the *Help-Giving Self-Concept Assumption*. According to this assumption, students with low help-giving self-concept are less likely to participate in teachable agent because they are unable to compose elaborated responses in a short time. So, elaboration support will particularly benefit the students with low help-giving self-concept. Based on the assumptions, low math self-concept may not influence the help-giving practice with the agent for the following reasons. First, low math self-concept may not influence this platform's participation as there is no fear of judgment since it is an agent. Second, we also provide the solution script of the problems, which the students can use to explain the solution to the agent. Third, the teachable agent task requires the students to explain the problem's steps to the agent. So, giving them elaboration support may improve students' help-giving skills and allow them to practice elaboration with the agent.

Overall Adaptive Collaborative Support Selection We described the design of the adaptive collaborative support considering both the platform the students are giving help in and their individual characteristics to select the appropriate support to display. Figure 13 demonstrates the high-level steps for this: the computational model gives a likelihood of students' constructive participation. This likelihood of constructive participation for a given platform was calculated by using the model equation described in Chapter 4 (Section 4.2.4). If the model outputs some likelihood of constructive participation, we display transactive support to the student. This support is designed to improve the ongoing collaboration by building on others' contributions. When the model outputs the opposite, the collaborative support system identifies which platform to give support to and then chooses the respective individual characteristic to display the support to the students. The algorithm following Figure 13 is given in Algorithm 1.

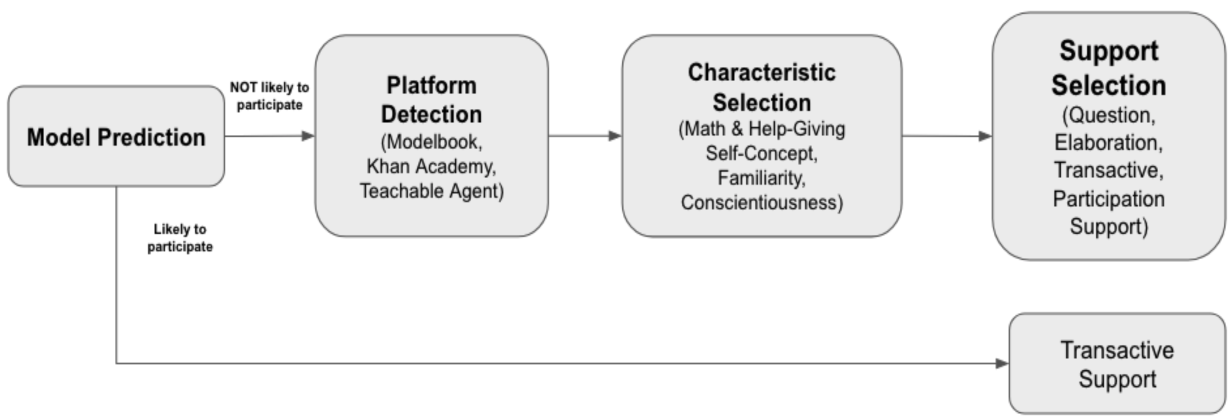


Figure 13: Selection of support based on platform and then students' individual characteristics.

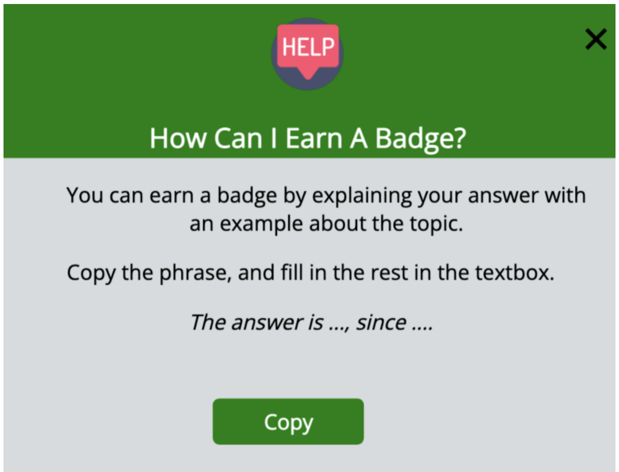
The primary adaptive collaborative support described in Chapter 5 (Section 5.3) only considered students' individual characteristics. However, the student characteristics seem to interact with the platform affordances and influence their collaboration. Hence, the revised adaptive collaborative support in this chapter aims to address the limitations of the previous approach and adapt the collaborative support depending on the platform the students are collaborating in. We investigated the effectiveness of this design in a classroom study (described later in this chapter).

6.2.2.2 Providing Badges as Positive Feedback

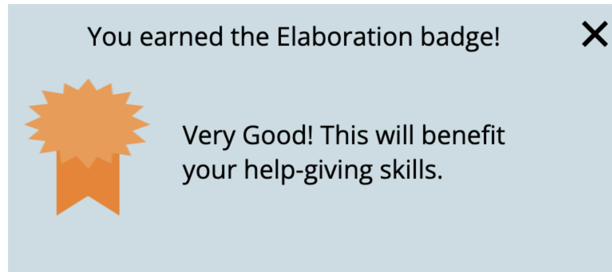
The second part of the adaptive support is the badges. We designed badges as a reward for student participation, i.e., a way to recognize and reinforce constructive participation behaviors. The selection of the badges was inspired by the peer interaction theory described in the previous chapter [121, 45]. There are nine badges (Table 22) for the four support groups mentioned above. The purpose of the elaboration badges is to encourage discussion. Question badges are designed to encourage asking clarification or thought-provoking questions. Transactive badges are expected to facilitate building on others' comments. Finally, the participation badges encourage the social aspect of collaboration. Following the positive feedback-only design, the badges included a content-free affirmative message (helps to reduce student uncertainty about the correctness of their participation) and a content-full message (helps to indicate the step taken by the student is productive for the conversation) [77]. An example is shown in Figure 14b.

Table 22: Definition of different badges used in Classroom Study.

Support Type	Badge Names (Definition)
Elaboration	Brainstorm (initiate a problem solution) Elaboration (explanation of a step to solve a problem) Summarize (briefly mention the main points of a solution)
Question	Question (wh questions)
Transactive	Add On (provide additional comments on a solution) Feedback (react to others' solution) Reflection (Providing own opinion on a solution)
Participation	Appreciate (recognize effort of others to solve a problem) Social (thanking fellow collaborators)



(a) Example of Elaboration Prompt



(b) Example of Elaboration Badge

Figure 14: Example of Adaptive Collaborative Support

6.2.3 Collaborative Support Implementation

As mentioned, we provided collaborative support in two ways across the platforms: a) encourage students to earn a badge via a prompt including a sentence starter to facilitate collaboration (Figure 14a); and b) reward badges designed as positive feedback based on students' utterances (Figure 14b). We now describe the implementation details regarding the adaptive support.

6.2.3.1 Displaying Support Prior/During Collaboration

Based on a student's self-reported motivation, a specific prompt was displayed at the beginning of each digital activity following Algorithm 1 (algorithm implemented following the high-level Figure 13). Figure 14a shows the interface of the prompt designed to encourage students to earn a particular badge by including a message and a sentence starter to help them participate. We expect this will encourage the students to get an elaboration badge. There was also an on-demand option within the interface to view the support as needed.

6.2.3.2 Classification of Student Utterances and Rewarding Badges

In the design study, we used a simple keyword-based approach to reward badges to the student based on their contribution to the discussion. In this revised design, we took a more automatic approach to detect specific utterances and reward badges as needed.

Different methods can be seen to automatically classify student utterances and provide feedback or context-sensitive collaborative learning support [94] as needed. For example, some studies use a keyword-based approach combined with a rule-based approach to detect certain utterances [114]. Another approach widely followed is the similarity measures [3]. This measure compares the given utterance with some pre-defined content and determines how closely the utterance match. Recent studies are using automatic detection of students' collaborative utterances using machine learning classifiers [94]. For example, Rosé et al. [94] applied text classification technology to a large CSCL corpus and identified the relevant linguistic features within a collaborative discourse. Joshi et al. [60] automatically assessed students' collaborative discourse using the concept of transactivity to summarize educational dialogue. Ai et al. [6] investigated transactive contribution in whole-group classroom discussions. Some research uses speech act classification methods to classify student utterances into a set of pre-defined categories, e.g., question or fact [96]. Samei et al. [96] proposed supervised machine learning models to explore the role of context (i.e., speech acts of previous utterances) for the classification.

In our case, we used semi-automatic detection of student utterances and rewarded a badge to reinforce the positive aspects of a student's utterance to guide student collaboration. We assessed students' utterances across the three platforms using an off-the-shelf classifier DialogTag and a custom-built keyword-based algorithm. "DialogTag" uses a neural architecture based on BERT to assign a dialog act to a sentence. DialogTag uses a subset of the Switchboard-1 corpus, which consists of about 2,400 two-sided telephone conversations among 543 speakers with about 70 provided conversation topics. We used this classifier to automatically classify student utterances into existing labels such as a question, feedback, reflection, appreciate, and social. Table 23 gives the example of the tags assigned by DialogTag and the badges we assigned based on that.

Table 23: DialogTag tags associated with different badges and their examples from the data.

Badges	Tags by DialogTag	Example from data
Question	Wh-Question, Open-Question, Rhetorical-Question, or Yes-No-Question	*Is there a different way to do this with non-proportional relationships?*
Feedback	Action-directive	*Try watching it again and take notes on what you do understand.:)*
Reflection	Agree/Accept	*yes i agree because the ratio 3:1 is in every paint splash painting*
Appreciation	Appreciate	No data
Social	Conventional-closing	*yes you're Welcome <name>..*

The keyword-based approach classified the utterances into one of the remaining four categories (brainstorm, elaboration, summarize, and add-on). For instance, *in addition*, *furthermore*, *moreover*, keywords were used to classify 'Add on' utterances. When any student fulfills the condition for a badge, s/he receives the notification about the badge immediately on the screen, as shown in Figure 14b. Students can view all the badges awarded to them for each platform on a badge page, which also shows the badges which have not yet been achieved.

6.3 Classroom Study: Implementation of the Revised Interactive Persona Tool and the Revised Adaptive Collaborative Support

We integrated the revised interactive tool and the revised adaptive collaborative support and deployed the system in a classroom study (n=76). The primary goal of the study was to examine the effectiveness of the interactive persona tool and the adaptive support provided to the students. In this section, we describe the classroom study and report the results related to *RQ3: How can we assess student motivation dynamically and within context?* and *RQ4: How can we design adaptive collaborative support using student motivation and context?* The findings of *RQ3* will provide us with insights into how students' motivation fluctuates during collaborative activities and how it influences students' help-giving participation. Similarly, the findings of *RQ4* will enable us to develop a collaborative support system that can adapt and respond to individual student's preferences and motivational states.

6.3.1 Method

The study was conducted across five classes in a middle school math classroom in the southwestern United States. We had two conditions: *limited technology* and *full technology*. Two of these classes were part of the limited technology condition, which used ModelBook as a source for the curriculum and to collaborate with others. The remaining three classes were part of the full technology condition, where the students used ModelBook, Khan Academy, and the Teachable Agent for collaboration. The students in this condition received support tailored to their individual characteristics and the specific platform they were in during their collaborative activities (described in Section 6.2). The individual characteristics were initially collected using a set of surveys/questionnaires at the beginning of the study. The students used the revised interactive persona tool throughout the study to modify their individual characteristics. In the full technology condition, we introduced an additional manipulation that involved utilizing individual characteristics to determine the type of adaptive support to display to the students.

Experimental Manipulation. Students from the full technology condition were randomly assigned to two conditions: *Survey Based Support* and *Persona Based Support*. In the *Survey Based Support* condition, the characteristic values were used from the survey which the students completed at the beginning of the study week. In this case, the support displayed to the students remained the same throughout the study because the characteristics obtained from the survey were fixed. In contrast, in the *Persona Based Support* condition, the characteristic values were used from the interactive persona tool, which the students used before each digital collaborative activity during the study. In this case, the support displayed to these students *might* have changed if the student changed their characteristics using the tool during the study. The students did not know which conditions they were assigned to, and all the students used the interactive persona tool to report any change in their motivation. Our hypothesis was that the students in the *Persona Based Support* condition would collaborate better as they received support based on their reported characteristics compared to the students in the *Survey Based Support* condition, where their characteristics are the same throughout the study.

This experimental design was a between-subjects design where each student was assigned only one condition; however, it is a within-class design as there might be multiple conditions present in the three classrooms.

6.3.1.1 Participants

All 135 students in the five classes were invited to participate, and 76 students provided parental consent. The breakdown of the number of students between the limited condition is 24, and the full technology condition is 52 (Survey Based Support condition=24, Persona Based Support condition=28)). Students ranged from 12 to 14 years old. Self-report demographic data for the 76 consenting students follows: 21 males; 36 females; 2 other, 1 prefer not to answer; 16 no response; 1 Asian/Pacific Islander, 5 Black or African American, 19 Hispanic or Latino, 1 Native American or American Indian, 17 White, 14 Other, 3 Prefer not to answer, 16 no response.

6.3.1.2 Measurements

Learning Assessments. We assessed learning outcomes using two isomorphic forms for the pretest and posttest. We intended to counterbalance the forms (i.e., half the students received Form A for the pretest, and half received Form B). However, during the study, students in one condition received Form A first and B second, and students in the other condition received Form B first and A second. Students received the opposite form for the posttest. Assessments were created in a collaborative process between the classroom teacher, modeling teacher, and research team to align the assessments with the content and state standards.

Motivation Pre-Measure. We asked the students to fill up motivational assessments twice: one prior to the intervention week and one after the students reached the end of the study. For the pre-motivational assessment, we used a survey consisting of the following: Math self-concept was adapted from the Programme for International Student Assessment [88], e.g., *'I have always believed that mathematics is one of my favorite subjects.'* Help-giving self-concept was adapted from Expectancy Value Theory [13] and Self Description Questionnaire-II [71], e.g., *' I'm confident that I can help others to learn the basic concepts taught in my math class.'* Familiarity was adapted from the Social Anxiety Scale for Children [64], e.g., *'I don't get nervous when I talk to new kids.'* Conscientiousness was adapted from the Big Five Questionnaire-Children [12], e.g., *'When I finish my homework, I check it many times to make sure I did it correctly.'* All the items were measured using a Likert scale ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree").

6.3.1.3 Procedure

Our partner school followed a block schedule where classes were longer than standard (each class was around 115 minutes) and met every other day. Each Wednesday, the school had early dismissal, so classes on that day were 60 minutes. The study ran over a four-week period with approximately 12.5 hours of instruction (7 class periods) for the full and limited technology conditions.

The study was conducted in Spring 2021, during the Covid-19 pandemic, thus, students had the option of attending class in-person or remotely. Remote students (n=34; 9 limited technology, 25 full technology) attended via Google Meet. Two teachers facilitated the student interaction during this period: a former teacher led the class online (since the school was closed to visitors), and a classroom teacher helped to conduct each class in person. The model teacher was the lead teacher for each session, and the classroom teacher was primarily in charge of classroom management.

There were seven class meetings over the course of four weeks. Students across both conditions followed similar curricula. On the first day, students were given an overview of the study, took the pretest (administered using Google Forms), completed a survey where they shared their demographic information and answered the motivational surveys, and were introduced to either the limited technology or full technology versions of ModelBook. The students in the full technology condition engaged in multiple digital activities (e.g., Modelbook, Khan Academy, and a Teachable Agent) where they worked in small groups, individually, or together as a whole class. On the second day, the students worked with ModelBook and were introduced to the first modeling activity (Perfect Purple Paint), which required students to find a ratio of red and blue paint to make a specific color of purple paint. In the third class, students discussed the definition of a ratio as a class and prepared an online whiteboard containing their small group's answers using the Modelbook chat feature. On the fourth day, students worked with the Teachable Agent individually for the first half of the class. Following this activity, they worked in their collaborative Modelbook chat groups to discuss an activity where they used the Paint Splash Phet tool to model their understanding of the material from class. On the fifth day, students collaborated in ModelBook's gallery discussion to chat about equivalent ratios and watched a video from Khan Academy on proportional relationships. On the sixth day, students were instructed to post a comment in Khan Academy after watching a video on ratios. After watching the Khan Academy video, the students used the Modelbook chat to post a question or answer questions posted by others. They also worked in their Modelbook chat groups to discuss real-life examples of proportions (e.g., cooking, baking, shopping) and started the second activity, Buggy Lab, an activity that allowed them to use their understanding of ratios and proportions to model the speed of a moving car. Finally, on the seventh day, students were instructed to find the question or comment they posted in Khan Academy to see if anyone had responded to their post. Students were then asked to post an additional response to that. Students also engaged in a gallery discussion to review their previous work on the Buggy Lab activity and took the post-test on their final day. In summary, after the completion of the study, we had one teachable agent interaction, two Modelbook digital discussions, and two Khan Academy discussions. Due to the hybrid arrangement, all the collaborative activities happened digitally.

6.3.2 Results

We first present the overall learning outcomes across the three conditions. Then we present the results related to the two research questions *RQ3* and *RQ4*. The findings of *RQ3: How can we assess student motivation dynamically and within context?* will provide insights into how students' motivation fluctuates during collaborative activities and how it influences their help-giving participation. Similarly, the findings of *RQ4: How can we design adaptive collaborative support using student motivation and context?* will enable us to develop a collaborative support system that can adapt and respond to individual students' preferences and motivational states.

Learning Outcomes. We began our analysis by looking into the learning outcomes for the three conditions *Limited Technology* condition, *Survey Based Support* condition, and *Persona Based Support* condition. We calculated adjusted learning gain using the formula *(Posttest-Pretest)/(100-Pretest)*. We used adjusted learning gain to measure student learning because it accounts for students' initial knowledge, providing a more accurate measure of their actual learning progress during the intervention, while post-test scores alone may not consider the starting point, leading to potentially misleading conclusions about the effectiveness of the intervention. We then conducted a one-way ANOVA to investigate the difference between learning among the conditions. Table 24 shows the mean and standard deviation of the learning outcomes for each condition. The one-way ANOVA revealed no statistically significant difference in learning among the three groups ($F(2, 57) = 2.311$, $p = .108$). This demonstrates that there is no evidence of learning in this study. This was possibly due to the limited opportunities for practicing help-giving, or the test did not reflect the skills we wanted to measure.

Table 24: Learning outcome scores for Classroom Study. N=# of Students.

Conditions (N)	Mean	SD
Limited Technology (N=20)	-.0075	.036
Survey Based Support (N=18)	-.0004	.036
Persona Based Support (N=22)	.0167	.039

6.3.2.1 Investigation of *RQ3:* How can we assess student motivation dynamically and within context?

The third research question aims to investigate the usage of interactive persona tool. We answer both *RQ3a: How do students use the persona tool?* and *RQ3b: How does interaction with the persona tool relate to students' help-giving participation across the different platforms?* using the data from the classroom study.

RQ3a: How do students use the persona tool?

Based on the pre-motivational survey responses, students were mapped with one of the 16 personas (100% match). The following are the total number of high and low students for each factor: MSC (21, 31), HSC (20, 32), FAM (32, 20), and CON (39, 13). To answer *RQ3a*, we investigated how students modified their personas throughout the six opportunities. Out of the 52 students, three students did not use the interactive persona tool. For the remaining 49 students, we break down the findings as follows:

Did students modify their initial personas at the first opportunity? For each student, the system-assigned initial personas are a 100% match with the questionnaire. Thus, any changes in one of the four factors would move the students away from their survey responses. We observe the following: (a) 33 students did not modify their initial persona; (b) 16 students modified their initial persona.

How did the students modify their initial personas at the first opportunity? The modifications done by the 16 students can be described as follows: (a) The average number of modifications per student is 1.25 (SD=0.45); (b) Students modified each of the factors in both directions (e.g., high to low or low to high).

How did the students modify their persona across the six opportunities? Students had six opportunities to use the tool before each collaborative activity. During these six opportunities, we observed the following (out of the 49 students): (a) 25 students did not change their persona at all and remained with their initial persona. This possibly means that for these students, the questionnaire and the initial persona indicated the same constructs; (b) 11 of them only changed their initial persona and remained with the same persona throughout. This suggests that for these students, the questionnaire and the initial personas targeted different constructs, and (c) 13 students modified their initial persona either at the later opportunities or modified both at the first opportunity and later opportunities (indicating multiple modifications). Similar to the design study, unidirectional and bidirectional modifications were observed on the same factors. This possibly indicates that students' motivation changed depending on the context or their interaction with others.

RQ3b: How does interaction with the persona tool relate to students' help-giving participation across the different platforms?

To answer *RQ3b*, we looked into students' modification of personas and their help-giving participation across the three platforms: Modelbook, Khan Academy, and Teachable Agent. We considered three sets of personas for this analysis: **Initial Persona** is the persona assigned to the students based on their survey responses, and **Modified Persona** is the persona when the student modified their initial persona the first time. **Final Persona** is the persona reported as of the last day of the students' using the persona.

We ran three linear regression models to predict students' total collaborative participation using students' (n=49) persona data as independent variables, i.e., initial, modified, and final persona, respectively, and the total collaborative participation as the dependent variable. A student's total collaborative participation on a platform is defined as the total number of utterances made by each student on that particular platform.

All three models are repeated for each of the three platforms. We would expect the modified persona will be more predictive than the initial persona as the students changed one or more motivational factors with respect to the surveys. However, the model with the final personas may behave differently as it includes students' later or multiple modifications during the study and thus may lack the ability to explain their total participation.

According to Table 25, For both Modelbook and KhanAcademy, the modified models give better and significant R^2 values. The modified persona regression models suggest that at least 25% of the variance in the total participation can be explained by the predictor variables for these two platforms. This could mean that the modified personas are better predictors, and personas might better assess student motivation for people who modified once at the first opportunity. However, for teachable agent, none of the models or factors are significant. It could be because students may be influenced by factors such as rapport

Table 25: Linear Regression Models (n=49) with different Personas as predictors and Total Participation as dependent variable. Here, MSC=Math Self-Concept, HSC=Help-giving Self-concept, Fam=Familiarity, Con=Conscientiousness. */** indicates a significant p-value.

Platform	**Model**	**R^2 (p-value)**	**Standardized Coefficients**			
			MSC	**HSC**	**FAM**	**CON**
Modelbook	initial	0.224 (.047)	.20	.27	.05	.31**
	modified	0.383 (.001)	.51**	-.08	0.06	.38**
	final	0.295 (0.01)	.45**	.03	.08	.27**
Khan Academy	initial	0.232 (0.025)	.34**	0.09	0.26*	0.10
	modified	0.259 (0.014)	0.31*	-0.10	0.39**	0.13
	final	0.182 (0.072)	0.33**	0.02	0.25*	.06
Teachable Agent	initial	0.133 (0.302)	.24	0.10	.20	-0.14
	modified	0.101 (0.459)	0.16	0.00	.27	-0.08
	final	0.135 (0.295)	.21	.07	.298	-0.05

with the agent, the perceived expertise of the agent, and the quality of the agent's feedback. These factors may have a stronger influence on students' help-giving participation compared to their help-giving motivation.

At the end of the study, the students completed a post-motivational survey where they were asked a single question *'What did you think about your profile page in Modelbook?'* Students' responses were in favor of the observed results. For example, one student said *"I think my profile page matched me and the way I feel about math."* (P27) which shows why many students probably chose not to modify the persona narrative. Another student said, *"... profile page fits me because ... i don't talk to that much but during the course of this model book, i started becoming more open to other people."* (P29) demonstrating the dynamic aspect of the student's motivation.

Discussion We described the design of the interactive persona tool that is contextually embedded to assess student motivation in intervals and presented the results of how the students used it to report their motivation. The design of the interactive tool is inspired by the features contextually-embedded, dynamic, and agentic. By embedding the tool and providing instructions to use it before each digital collaborative activity within the digital textbook, Modelbook, the entire experience became contextually integrated and seamless for the students. Moreover, by providing the students with a broader range of personas (16 personas compared to 4 personas in Design Study in Chapter 5) to choose from and the ability to modify their assigned persona multiple times, we ensured that they had ample opportunities to actively engage with and adjust their motivational profiles within the tool. This revision was crucial for investigating the dynamic assessment of motivation in different contextual settings, as it allowed us to explore how students' motivation varied and responded to changes in their learning environments. Lastly, providing students with the option to modify their motivation or keep it as is, the ability to see the updated narrative with the latest modified response and change again as they needed promoted student agency. This acknowledged their role as active participants in the learning process.

The results of *RQ3* demonstrated the utility of this interactive tool. We observed nearly half of the students did not modify their initial persona at all. For these students, this indicates the questionnaire measured the same motivational constructs as the initial persona, and their motivation remained constant throughout the study. For the other students, who modified their motivation multiple times, when viewing the survey items in a narrative, the students reflected on their motivation at that moment. This is consistent with the literature. For example, Orji et al. [82] suggest students' motivational states are likely to vary during learning interactions. Moreover, the regression models demonstrated that the modified personas were better predictors to explain the total participation across different platforms. It indicates that the personas are a better assessment of student motivation for people who modify once at the first opportunity. These findings indicate that the interactive persona tool could be an alternative way to assess student motivation within a collaborative context.

In summary, the interactive persona tool is a step toward understanding the dynamic assessment of motivation within the ITSs context. According to [37], the focus on dynamic adaptation to student motivation using ITSs is increasing. Excluding studies that used surveys for measuring motivation, research modeling students' motivational states are small, suggesting a wider exploration of techniques for inferring the motivational states of students [82]. Also, this way of assessing motivation might be particularly useful for developing ACLS. We included multiple motivational factors as part of our personas as we observed a combination of different motivational factors, rather than a single construct, was needed to describe student participation. In motivation research, the multidimensional perspective of motivation suggests motivation cannot be adequately explained in terms of a single construct [72]. On a practical level, it may be intractable for ACLS to respond differently to permutations of multiple interacting motivational factors. Thus, leveraging personas can be a way for ACLS to prioritize interventions based on logical clusters of individual characteristics.

6.3.2.2 Investigation of *RQ4:* How can we design adaptive collaborative support using student motivation and context?

The fourth research question aims to investigate the usage of adaptive support. The sub-research questions are *RQ4a: How did the students use the support (sentence starters and badges) across the three platforms?* and *RQ4b: What is the relationship between student support usage and student help-giving participation?*

We utilized data from the full technology condition to address these sub-research questions. It is important to mention that we had a manipulation (explained in Section 6.3.1) within the full technology condition, involving the survey-based condition and the persona-based condition. This manipulation aimed to investigate how modifications of students' motivation in the persona-based condition influenced the support they received and whether it impacted their collaborative participation. We hypothesized that students in the persona-based condition would collaborate more effectively and, consequently, learn better than those in the survey-based condition. However, we did not observe significant differences between the two conditions when we analyzed students' collaborative participation (i.e., minimal, facilitative, constructive) across the three platforms. One possible reason for this finding is the mismatch between a student's modification of a characteristic and the factor associated with that platform for displaying support. For instance, students were asked to modify their motivation before each digital collaborative activity. If a student modified the help-giving self-concept characteristic and the next activity was in Modelbook, the support displayed in Modelbook would not change, as the priority was given to the math self-concept. The support would be based on the high/low value of the math self-concept for that student in the context of the Modelbook. As a result, the analysis did not yield conclusive results, and we chose to report the findings of *RQ4* sub-research questions by collapsing these two conditions.

Before diving into the sub-research questions, we look into the different types of support the students received. Out of the 52 students, 65.38% received only transactive support, 19.23% received two different supports, 9.62% received three different supports, and 5.77% did not receive any support at all. The reason for the transaction support to be the predominant support could be explained using the Algorithm 1 (see Figure 13). It could be due to the likelihood of participation being greater than 0.3 (based on previous data) or a large number of students reporting high conscientiousness (out of the 52 students, 39 students reported high conscientiousness via the survey). Due to this issue of having one support prevalent over the others, we shifted our focus to analyze how students used the support irrespective of the support they received.

RQ4a: How did the students use the support (sentence starters and badges) across the three platforms?

Sentence Starters Out of the 52 students, 30.76% of them used sentence starters across the three platforms (M=4.34 starters per student, SD=4.71), and the remaining students did not. Of the total sentence starters used by the students, 88.5% are applied during Modelbook collaboration, and the rest are used during Khan Academy participation. Students did not use any sentence starters while using the teachable agent.

Badges Out of the 52 students, 82.69% of them earned at least one or more badges across the three platforms (M=4.16 badges, SD=4.59). Among the total badges earned, 26.42% were earned using sentence starters, e.g., *"In addition you can see that their graph stayed a straight line the entire time."* earned an Add on badge as 'In addition..' sentence starter was used. Utterances such as *"rate of change on a graph means slope, and specifically for this graph the rate of change is 1:10 since that's the change in y/x"* earned an elaboration badge without using any sentence starters. There are cases when students did not use the sentence starters while participating constructively and did not earn any badges. It could be because while giving definitions, describing an activity, or comparing their groups' work with other groups, their responses might have contained very specific domain-related words, e.g., in response to an activity related to ratios, one student replied, *"the ratios are equivalent which means that the paint will be the same color but will be in different quantities."*

Furthermore, of the total badges, 50%, 12.26%, and 37.74% were awarded in the Modelbook (MB), Khan Academy (KA), and the Teachable Agent (TA) platform, respectively. We calculated the percentage of badges earned on each digital activity and took a deeper look following the order of the activities that happened (Table 26):

Table 26: Percentage of different badges earned in different digital activities.

Badge Name	TA (%)	MB-1 (%)	KA-1 (%)	MB-2 (%)	KA-2 (%)
Elaboration	57.5	29.87	7.14	44.83	50
Brainstorm	27.5	5.19		37.93	8.33
Summarize					
Question	10	24.68	85.71	10.34	16.67
Add On	2.5	10.39			16.67
Feedback	1.25	3.9	7.14	3.45	
Reflection		19.48		3.45	8.33
Appreciate		2.6			
Social	1.25	3.9			

First, earning badges differed across the platforms, probably due to platform or task affordances, e.g., the appreciate, or the social badges may not be appropriate for Khan Academy as it is a question-answer-based platform. The summarization badge was not earned on any of the platforms, indicating that with the tasks to be completed, providing a summarization was not common. Rather, the elaboration badge may have served a similar purpose. Second, earning badges differed within the same platform itself. The two activities at Khan Academy were at least two weeks apart. During the first activity, students mostly asked questions, but in the second activity, there was a diverse range of badges earned by the students, possibly because they participated in the same topic and replied to others instead of asking new questions.

Validity of the badges We examined whether the badges earned by the students were valid. Out of the total badges, 78.77% were correctly assigned. For the rest of the badges, we observed the following reasons for incorrect assignment: misusing the sentence starters (13.68%); not having content-related info (4.72%); and incorrect assignment (2.83%), e.g., an utterance was awarded a reflection badge, but an elaboration badge would have been more appropriate.

RQ4b. What is the relationship between student support usage and student participation?

In order to examine the relationship between support and students' constructive participation across the three platforms, we ran three separate regressions. First, we ran a simple linear regression with the average number of constructive participation in Modelbook as a dependent variable and the average number of sentence starters used in Modelbook, and the average number of badges earned in Modelbook as independent variables, controlling for motivational factors (MSC, HSC, CON, FAM) and pretest.

Table 27: Simple Linear Regression with Average No. of Support as predictors and Avg. No. of Constructive Participation as the dependent variables controlling for motivational factors and pretest. Here, MSC=Math Self-Concept, HSC=Help-giving Self-Concept, Fam=Familiarity, Con=Conscientiousness. Significant predictors are indicated in bold.

Predictors	Modelbook	Khan Academy	Teachable Agent
R^2	**0.792 (<.001)**	**0.466 (.001)**	0.222 (.238)
MSC	-0.012 (.904)	- 0.015 (.920)	0.149(.464)
HSC	-0.093 (.353)	0.027 (848)	-0.024(.912)
Fam	0.115 (.178)	**0.391 (.005)**	0.070(.694)
Con	0.091 (.291)	0.219 (.099)	0.155(.402)
Pretest	**0.273 (.016)**	**0.381 (.007)**	0.219(.244)
Avg. No. Sentence Starters	**-1.165 (<.001)**	-0.126(.355)	0
Avg. No. Badges	**1.449 (<.001)**	-0.153(.270)	0.297(.155)

Table 27 indicates that both Modelbook (R^2=.792, F(7,39)=17.418, p<.001) and Khan

Academy (R^2=.466, F(7,43)=4.492, p=.001) models are significant. In Modelbook, pretest (β=0.273, t=2.552, p=.016), and the average number of sentence starters (β=-1.165, t=-6.524, p<.001) and badges (β=1.449, t=7.734, p=p<.001) being the significant predictors in the model. In the case of Khan Academy, the sentence starters and badges are not predictive; instead, pretest (β=0.381, t=2.841, p=.007) and familiarity (β=.391, t=3.021, p=.005) are the significant predictors in the model. For the teachable agent, the model is not significant R^2=.222, F(6,36)=1.425, p=.238, and there is no significant predictors.

We conducted a regression using posttest scores as the dependent variable and pretest and average badges earned in each platform as independent variables. The model is significant R^2=.706, F(4,28)=14.384, p<.001, however, only the pretest is the significant predictor (β=.7, t=5.442, p<.001). The average number of badges earned on different platforms did not influence student learning.

Discussion Our findings related to *RQ4* particularly highlight how the support system that imperfectly models the domain might still form the basis for feedback that leads to better collaboration in ill-defined collaborative ITSs. Leveraging student motivation and the platforms, the collaborative support was designed using a combination of sentence starters and badges. We summarize the findings: first, badges did not have a significant influence on learning outcomes. The lack of learning could be due to the voluntary nature of student participation and the limited time to engage with the system [103]. This finding is consistent with previous research that did not find any negative effects on learning that could be tied to badges [47]. In our case, we designed the badges following a positive-feedback-only design to reinforce the positive aspects of a student's utterance, e.g., giving help using explanations and asking clarification questions. Second, some students used sentence starters which helped them to participate and earn badges. In Modelbook, both sentence starters and badges were predictive for constructive participation. However, a negative coefficient for the average number of sentence starters used indicates students who used sentence starters less frequently were more likely to participate constructively. This suggests that the use of sentence starters to facilitate help-giving may not be ideal for synchronous chat-like discussions. Moreover, the findings show that the badges were correlated with constructive participation independent of the pretest in Modelbook, suggesting that it is possible they may have led to more constructive participation in that platform. Students also had positive feedback on the badges. In a post-motivational survey, 30 students answered a single open-ended question on their perceived usefulness of the badge system. Students' responses (enclosed in parenthesis) revealed that badges were a useful way to guide them during collaboration (*"I think the badges are a good way of showing my help was useful to others."*), enhance motivation by providing a sense of achievement (*"I thought the badges were well received for the tasks I did."*), and acknowledge their effort to help others (*"I like the badges since they do acknowledge what you should do."*). Lee et al. [66] describes a similar observation, i.e., how badges helped to increase students' sense of recognition, promoting their motivation and engagement. Third, we observed badges operate differently across different contexts, and in different contexts, different badges might be appropriate. Hence, to design such support, context, and task affordances should be considered. For example, students earned fewer badges at Khan Academy than on other platforms. This could be because authoring questions and responding to questions are more demanding and time-consuming tasks. Since the students were asked to complete the task within the classroom, some students likely didn't get enough time to respond and hence didn't get badges. Moreover, the teachable agent allows spoken synchronous collaboration, so choosing sentence starters in the middle of such activity could have impacted the flow of the conversation. Future studies should investigate how different contexts impact sentence starter and badge design and further explore the benefits of positive feedback in a cross-platform collaborative environment.

6.4 Conclusion

This chapter describes the revised interactive tool to assess students' motivation during the study. It also describes the revised collaborative support based on the student's individual characteristics and the platforms. We then describe a classroom study where we implemented both the revised interactive tool and the adaptive support. The results describe how the students used the tool to report their motivation prior to collaboration. The results also demonstrated the badges were successful in facilitating collaboration in one of the platforms. Future studies should investigate platform affordances while designing the support using badges.

7.0 Investigating the Generalizability and Transferability of the Computational Model

In Chapter 4, we presented the development of an explanatory model that aimed to describe how students with different characteristics participating in different platforms choose to participate constructively in a given platform. By leveraging the data collected from the design-based research study (described in Chapter 3), this modeling approach is able to explain why certain students exhibited constructive participation behavior while others did not. The need for such an explanatory model in education research is emerging as well. Having an explanatory model which gives insight into the learners, the learning process, or the instructional context can be useful. The model can help to gain insights into why certain learning behaviors occur, helping researchers and educators to make informed decisions and design more effective instructional strategies. In favor to this, Rose et al. [95] argue that machine learning predictive models alone are not always the answer, *explanatory learner models* can offer actionable insights that might advance both learning science and educational practice. Moreover, in the context of educational research, it is typical to have a small sample size due to various reasons, e.g., limited participant availability, consent issues, etc. In such a case, due to the lack of enough data, predictive models can be difficult to generate. All these encouraged us to focus on developing an explanatory model.

Although the primary objective of an explanatory analysis is to develop a model that explains the underlying mechanisms and factors influencing the observed outcomes, prioritizing understanding over predictive accuracy, it is crucial to report the model's predictive qualities alongside its explanatory power to ensure a comprehensive evaluation [100]. By including information about the model's predictive capabilities, researchers can assess its overall performance and make fair comparisons with other models, leading to a more thorough understanding of its effectiveness and practical applicability. Hence, in this chapter, we investigated whether the computational model is generalizable to other study contexts.

We explore the generalizability in two ways. First, we evaluate how well the computational model explains students' constructive participation in unseen data to assess its generalizability and predictive performance. By testing the model on new and previously unseen data, we can determine its ability to accurately explain students' participation behavior in different contexts, beyond the data used for model development. This evaluation helps to ensure that the model's explanatory power is not limited to the specific dataset used during training and that it can provide meaningful insights and predictions in formal classroom collaborations. So, we address one of the last two research questions of *RQ2*, *2c: How well does the computational model explain the students' constructive participation in unseen data?* Second, we explore whether the model is transferable to a new study context. Due to COVID, we had the opportunity to conduct studies in different setups, e.g., the design study in Chapter 5 was conducted online, and the classroom study in Chapter 6 was conducted in a hybrid setup. By transferring the model, we assess whether the model can be extended to incorporate a new parameter, such as in-person vs. online vs. hybrid setup, to improve its explanatory as well predictive performance. By doing so, we can determine if the model is robust enough to consider various contexts beyond the combined factors used during its development, thereby enhancing its applicability and generalizability in different educational settings. We address the last sub-research question *2d: Is the computational model transferrable beyond the context it was built?*

This chapter has two main sub-sections: in Section 7.1, we explore how well the model explains unseen data. In Section 7.2., we describe an extension of the model beyond the context it was developed.

7.1 Exploring the Computational Model with Unseen Data

To examine the generalizability, we apply the computational model developed in Chapter 4 on new datasets obtained from the subsequent studies (brief summary of the datasets in Section 7.1.1). We measured the accuracy of the predicted probabilities by calculating Brier Score. Brier Score measures the overall accuracy and calibration of the predicted probabilities, where a lower score indicates better performance. We explained Brier Score in detail in Chapter 4 in Section 4.3.1. Since the comparison of the different models are done using different datasets, the models will not be comparable by means of likelihood-based criteria such as AIC/BIC, so we will be using Brier Skill Score instead.

7.1.1 Description of the Datasets

Description of the Datasets We have student participation datasets from three different studies: design-based research study (Chapter 3), design study (Chapter 5), and classroom study (Chapter 6). Each of these datasets includes student participation data from all three platforms: Modelbook digital discussions, Khan Academy postings, and Teachable Agent problems. We used the Help-Giving Coding Scheme (section 3.3.2.1) to label each interaction as constructive participation or not. Table 28 shows the collaboration mode, the total number of observations, and a breakdown of the number of observations across the platforms for each study. We noticed that the Modelbook has the highest number of observations for all the studies. For the remaining part of this chapter, we refer to each dataset with respect to the collaboration mode.

Table 28: Description of dataset for each study. Here n is the number of students.

Study (n)	Collaboration Mode	Total Number of Obs.	No. of Obs. in Modelbook	No. of Obs. in Khan Academy	No. of Obs. in Teachable Agent
Design Based Research (n=16)	In-person	755	462	148	145
Design Study (n=17)	Online	275	175	22	78
Classroom Study (n=52)	Hybrid	928	657	81	190

7.1.2 Experiments

We conducted two experiments where we used different datasets to build the model and investigated the sub-research question *2c: How well does the computational model explain the students' constructive participation in unseen data?*

Experiment 1: In the first experiment, we trained the model with the in-person dataset and investigated how the model is generalizable to the online and the hybrid study data separately. In this case, the model does not have any information about the online participation and we'd like to see how it behaves with the unseen online and hybrid data.

Experiment 2: In the first experiment, the model does not have any knowledge about the online data, and was applied to datasets that both had students' online participation. So, in the second experiment, we combined the in-person and online data, and use the combined dataset to build the model. We then investigated how this model is generalizable to the hybrid study data. Since, this model has knowledge about both in-person and online participation, our hypothesis is that the model from the second experiment might lead to a better fit for the hybrid dataset compared to the model in the first experiment.

7.1.3 Results

We report the results of the two experiments described above. We use Brier Scores to report the quality of probability predictions for each model and use Brier Skill scores to compare the predictions across the models. We don't use any likelihood scores such as BIC as they are used to compare models across the *same* dataset. We also chose the Brute Force Main Interaction model (initially described in Chapter 4) to make a comparison with the computational model. This is a model with a group of main and interaction effects between the individual characteristics (MSC, HSC, and Fam) and Platform Type. We chose this model because it uses the same explanatory variables as the computational model, but it has no theoretical constraints as with the computational model.

7.1.3.1 Results from Experiment 1

In this experiment, we trained the model using the in-person dataset. To examine whether the model is generalizable beyond the data that it was built, we applied the model to the other two datasets, online and hybrid data.

Table 29 reports the Brier Scores for the computational model and the Brute Force model. We added the Brier Scores for in-person data from Chapter 4 for comparison. We observe two things: first, for each dataset, the Brier Scores for the computational model are lower than the respective brute force model which leads to positive Brier Skill Scores

Table 29: Computational Model built on in-person data and generalization test on unseen Online and unseen Hybrid data. The bold marks the best result.

Model Name	Brier Score on In-person	Brier Score on Online (unseen)	Brier Score on Hybrid (unseen)
Computational Model	**.2**	**0.267**	**0.184**
Brute Force Model	0.207	0.282	0. 191

(Brier Skill Score = 1 - $\frac{BS}{BS_{ref}}$, detail in Chapter 4). A positive Brier Skill Score indicates the computational model provides more accurate predictions with respect to the reference model (here the reference model is the Brute Force model). Second, for the computational model, in terms of Brier Score, we observed that, the Brier Score on Hybrid data < Brier Score on In-person data < Brier Score on Online data. Based on this, the computational model seems to generalize well on the Hybrid dataset but not on the Online dataset. Similar behavior is observed for the brute force model as well.

7.1.3.2 Results from Experiment 2

In Experiment 2, we trained the model using the combined data of In-person and Online datasets. We then examined if the model built on the combined dataset is generalizable to the Hybrid dataset. Our hypothesis is that since this model has knowledge about both in-person and online participation, it might generalize (in terms of predicted probability) better than the model in Experiment 1. Similar to Experiment 1, we use the Brute Force model to compare the Brier Scores.

Table 30 presents the BIC scores and the Brier Scores on the Hybrid dataset for the computational model and the Brute Force model. The first row in the table is added from Experiment 1 for comparison. We do not include the BIC score in the first row since the model is built on a different dataset.

We use Brier Skill Score to compare the computational models built on two different

Table 30: Brier Scores for the models on Hybrid data.

Model Name	Training Data	BIC	Brier Score on Hybrid (unseen)
Computational Model	In person		0.184
Computational Model	In person + Online	**1317.87**	**0.187**
Brute Force Model	In person + Online	1366.24	0.195

datasets. Considering the model built on in-person data as the reference model, we get a Brier Skill Score -.016 (=1-.187/.184); a negative Brier Skill Score indicates the current model (computational model built on the combined data) performs poorly than the reference model. This finding is opposite to our proposed hypothesis.

Interestingly, comparing the computational model and the Brute Force model on the combined data, we observe the computational model has a lower BIC score, as well as a positive Brier Skill Score (considering the Brute Force as the reference model in this case), indicating the computational model performs well. This result is consistent with the findings from Experiment 1.

7.1.4 Discussion

Modeling collaboration is one of the important tasks within an ACLS. While most of the efforts in this area use student participation data to build a predictive model for collaboration on a single platform, we focus on developing an explanatory model to define students' constructive participation across the different platforms. In this section, we investigated whether the computational model is generalizable to unseen data.

In the first experiment, we built the computational model using the in-person dataset. We observed the computational model performs well compared to the Brute Force model in all three different datasets. In terms of predicting probabilities in the unseen data, we observed the computational model, as well as the brute force model, performs well for the

Hybrid data. This indicates that the model could possibly generalize well to unseen data. However, both models perform poorly on the online data. It could be due to the fact that online data has less number of observations or it could be that the data is noisy. Furthermore, from Table 29, the Brier Score for both the models is above 0.25. By definition, Brier Score above 0.25 indicates the model is predicting everything as 50% [1]. This further supports the fact that the performance issue could be due to the data. We recognize that the unseen data are from completely separate studies, and each differs in how they were conducted with different sets of students. Since the model performs well on the Hybrid dataset, this result is encouraging.

In the second experiment, we found our hypothesis that the computational model built using the combined dataset (in-person and online data) would perform better compared to the computational model built using only the in-person, does not hold. But the computational model itself performs better than that of the reference model (Brute Force model). Using the knowledge from Experiment 1, we could say it could be some noise within the Online data that is influencing model performance.

Overall the results seem encouraging that the computational model performs well compared to a reference model. In terms of generalization, the model also showed promise in explaining unseen data, although further investigation is needed.

7.2 Transferability of the Computational Model

Using the qualitative data based on the design-based research study and the literature, we developed the computational model in Figure 15. This study was held in person in a classroom setup. To recap, the model included two predictors, $F12\_combined$ and F_3 (described in Section 4.2.4). F12_combined sub-model is a combination of Help-Giving Self-Concept (H_p), Familiarity (F_p), and Synchronicity (σ_e). F_3 is a combination of math self-concept (M_p) and Platform Type.

As a part of this thesis, we conducted two more studies after the design-based research: a design study and a classroom study. Due to the COVID-19 pandemic, the design study

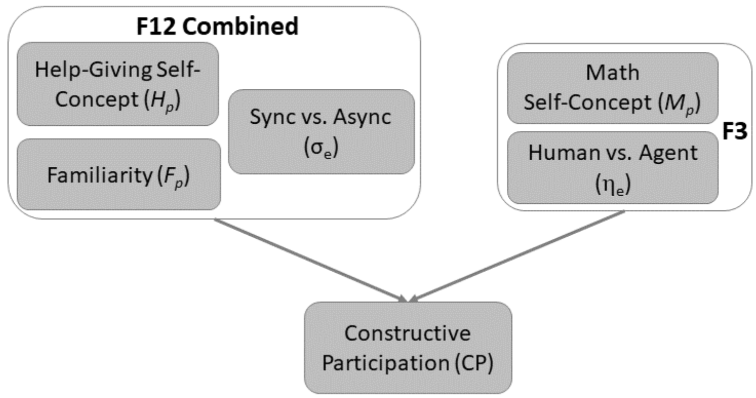


Figure 15: Constructive Participation Model Visualization

(described in Chapter 5) was conducted online. Almost six months later, the classroom study was conducted in a hybrid setup (i.e., a combination of in-person and online). This gave us an opportunity to explore whether the computational model developed in Chapter 4 is transferable beyond the original study.

In order to do this, we introduce a function F_4, which is a combination of the individual characteristic, conscientiousness (C_p), and a contextual variable, *presence* (γ_e). Conscientiousness refers to the dutifulness of a task. As with any other individual characteristic, C_p takes the values {0, 1} for {low, high}. The *presence* variable has two levels, {online, in-person}, and takes the value {0,1}. Similar to the other sub-models described in 4.2.3, we define F_4 by a two-way table (Table 31). Our reasoning about the corresponding two-way table follows the assumption that if a student has low conscientiousness, then the likelihood of constructive participation decreases in an online environment (therefore negative value), but may or may not participate constructively in an in-person setup, so the likelihood remains the same (therefore zero). In contrast, if a student has high conscientiousness, then the likelihood of constructive participation increases both in the online and in-person environment, but the likelihood might be more in an in-person setup, because the students are in a formal environment (with students and teachers around them). Quantities are thus assigned to the corresponding Table 31. Adding this F_4 parameter in the model will allow us to investigate whether our modeling approach is transferable to a new study context.

Table 31: F_4 Table with values filled in.

	γ_e = Online	γ_e =In-person
C_p=high	0.5	1
C_p=low	-1	0

Figure 16 presents a graphical representation of the extended model.

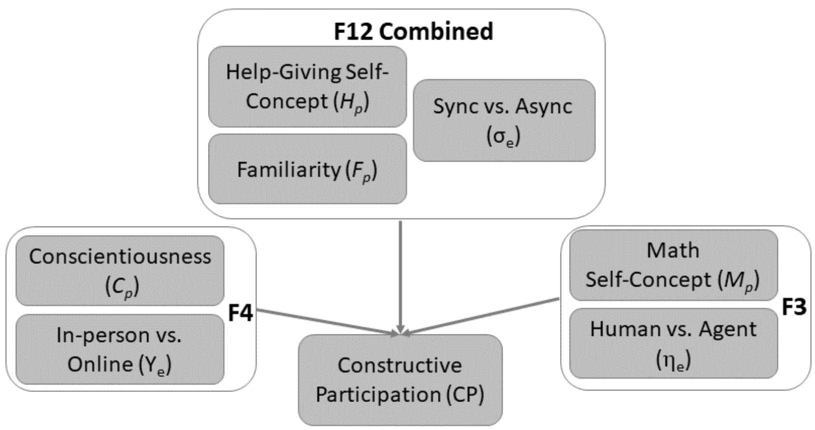


Figure 16: Computational Model with F_4 Parameter

7.2.1 Building the Extended Computational Model

The sub-research question we want to answer through this section is *2d: Is the computational model transferrable beyond the context it was built?* To explore the effects does F_4 sub-model on student constructive participation, we built the extended computational model (Figure 16) with the combination of in-person and online data and ran a GLM (general linear model) function and specify family="binomial" so that R fits a logistic regression model to the dataset.

7.2.2 Results

We get the odds ratio (exponentiated coefficients) and their confidence intervals using standard error in Table 32:

Table 32: Coefficients, Odds Ratio, and Confidence Interval for the Extended Computational Model. ** indicates p-value < 0.001, indicating significant predictors.

	Coefficients	Odds Ratio	2.5% CI	97.5% CI	p-value
Intercept	-.7274	0.4831227	0.3772750	0.6186668	0.00**
F12_combined	.3538	1.4244655	1.2652958	1.6036581	0.00**
F3	.6313	1.8801236	1.5006430	2.3555667	0.00**
F4	-.7742	0.4610558	0.3480287	0.6107902	0.00**

All of the factors (including the intercept term) presented to be significant predictors. If the odds value is less than 1.0, this indicates that an increase in the predictor decreases the odds of the outcome occurring. In contrast to F12_combined and F_3, each additional increase in F_4 is associated with a 53.9% decrease in the odds of students participating constructively.

Next, we investigate the BIC scores and Brier Scores for the extended computational model and Brute Force Model. We applied both models to the unseen Hybrid data and investigated the generalizability of the extended model. We recomputed the Brute Force model with the additional parameters (main effects and interaction effects of conscientiousness, presence). We report the scores in Table 33.

We see similar results as described in the previous section. The extended model has a lower BIC score indicating better model selection for generating the data. On the training data, the Brier Score is almost similar, but the lower Brier Score on Hybrid data indicating the extended model generalized to the unseen data. In addition, considering Brier Skill Score, the extended model performs well too, compared to the Brute Force model.

7.2.3 Discussion

In this subsection, we explored whether we can add a new parameter as a way to transfer the computational model across different study contexts. In order to do that, we introduced

Table 33: Extended Computational Model built on a combination of in-person+Online data and generalization test on unseen Hybrid data. The bold marks the lowest scores.

Model Name	BIC	Brier Score on In-person +Online	Brier Score on Hybrid (unseen)
Extended Model	**1295.93**	0.212	**0.205**
Brute Force Model	1361.57	**0.211**	0.232

the F_4 sub-model with the individual characteristic, Conscientiousness, and the contextual factor, Presence. The extended model seems to better fit the unseen data in terms of the Brier Score.

7.3 Conclusion

We followed up with the model developed in Chapter 4 in this chapter. We used the model to describe how well it generalized to unseen data. In this chapter, we also investigated whether our model is extendable with new parameters to a new study context.

8.0 Discussion

In this thesis, we describe the development of a learning environment, UbiCoS, which includes three digital learning platforms with different characteristics. For instance, the digital textbook, Modelbook allows synchronous text-based collaboration, Khan Academy promotes asynchronous collaboration by posting questions and providing answers to questions, and the teachable agent allows a safe space for practicing elaboration and explanation. Students collaborate on these different platforms, practice their help-giving skills, and hopefully transfer their skills from one platform to another. Our overall goal is to provide personalized collaborative support to middle school students across these platforms. In order to build such an adaptive system, we first need to understand how student collaboration varies across these platforms, which are inherently different. We also investigated which individual characteristics influence student collaboration in these platforms and leveraged those characteristics to model and design collaborative support. Based on the findings, these individual characteristics seemed to interact with the platform characteristics, and so we developed a contextually embedded tool that students can use to report their motivations prior to each digital collaborative activity. The remaining part of this section briefly summarizes each overarching research question, which helped build the single adaptive collaborative system using student motivation.

UbiCoS System and Student Help-Giving Behavior. Our first overarching research question *RQ1: How does student participation vary across the three different learning environments?* aims to understand the variation in student collaboration across the different platforms within UbiCoS (Chapter 3). We conducted design-based research in three cycles, where we deployed UbiCoS with incremental changes. We developed a three-level coding scheme following Webb's help-giving [120] theory and Berkowitz's transactivity [14] theory to examine students' mathematical reasoning skills or social skills of the students, which is where the collaborating students are expected to benefit [81]. By coding student utterances for an explanation, elaboration with examples, and transactivity, we focus on all the necessary skills for mathematical reasoning. We used this coding scheme to analyze student help-giving participation across the different learning platforms and found that the collaboration varies in accordance with the platform. For example, the constructive participation in Modelbook doubled in cycle three, while the quality of Khan Academy posts remained fairly stable throughout the three cycles. Additionally, the design of UbiCoS included multiple contexts that served different populations. For example, those students who were hesitant to participate in face-to-face whole-class discussions felt comfortable participating in the asynchronous or synchronous platforms. The same students participated differently across the different platforms, e.g., some students participated at a higher rate than average in Modelbook but lower than average in Khan Academy, and vice versa. Similar results are seen in [84], where Oztok et al. looked into student participation in both synchronous and asynchronous platforms in the *same* environment and reported the quality of messages in the two different platforms is different. The difference in help-giving participation across the platforms encouraged us to investigate the underlying factors that influence a student's willingness to participate in each platform. By identifying the specific motivations and tailoring support accordingly, we aimed to create an effective collaborative learning environment that foster constructive participation across all three platforms in UbiCoS.

Modeling Help-Giving. The second overarching research question we addressed is *RQ2: How can we design an explanatory model using student motivation and contextual factors to explain students' collaborative behavior?* In Chapter 4, we took a deeper dive into the varying levels of student participation and conducted a semi-structured interview using Expectancy-Value Theory. According to this theory, students are motivated to engage in help-giving when they perceive the value or importance of giving help on different platforms. Our goal was to understand the factors influencing student collaboration across the three different learning platforms. After analyzing the transcripts, we identified certain motivational characteristics that played a role in student help-giving. They are math self-concept, help-giving self-concept, familiarity, and conscientiousness. Interestingly, we observed that these characteristics differed depending on the specific platform. For instance, a student who was confident in providing help and supporting their friends in Modelbook mentioned that they did not value helping others at Khan Academy because they preferred assisting their friends rather than strangers. This finding inspired us to consider the student's motivation to collaborate and the platform they were collaborating in rather than focusing solely on motivational characteristics or solely on platform features. We followed an explanatory modeling approach, combining several individual and platform characteristics using qualitative data and theory and developing an explanatory model to explain students' constructive participation across the three platforms. In the context of ITSs, this explanatory model could be beneficial in understanding the underlying reasons for a particular outcome and providing personalized feedback. For instance, unlike a predictive model that could be used to identify students who are likely to struggle with constructive participation based on their past performance, an explanatory model can provide insights that the student might be struggling due to low motivation in an asynchronous platform. The system could then offer scaffolded support such as hints or explanations. The system could follow patterns of disengagement and offer timely reminders or motivational messages to keep students on track.

In Chapter 7, we investigated the generalizability of the computational model to unseen data. Additionally, we explored whether the model could be transferred to other study contexts beyond the ones used for its initial construction. By conducting these investigations, we aimed to evaluate the adaptability of the model, which is crucial for its practical application in different educational scenarios. The model exhibited generalization (in terms of predicted probability calculated using the Brier Score) to other study data. Additionally, when adding another parameter combining conscientiousness and presence, the model indicated a better selection (in terms of BIC score). However, there is still room for improvement in using this model in real-time and improving its generalizability. It is important to note that the evaluation data used to assess the model came from completely different studies involving students with potentially different motivations. We recognize that this is a hard problem with small data obtained within a short time period, with many variances. This research contributes to the broader field of educational technology by emphasizing the value of considering explanatory models with predictive power. By incorporating individual and contextual factors, this approach enables the design of effective interventions and support systems within collaborative learning environments.

Interactive Persona Tool. By now, we have established that motivation plays a key role in student participation in UbiCoS, and this motivation seemed to vary when students collaborate on different platforms. This change in motivation can be explained by the multifaceted nature of motivation in the literature. This line of research emphasizes the effect of contextual and environmental factors on motivation [72, 82]. Following this nature of motivation, we have our third overarching research question *RQ3: How can we assess student motivation dynamically and within context?*. To model and adapt support based on the dynamic nature of motivation, we believed this is an important aspect to investigate. In Chapter 5 and in Chapter 6, we describe the incremental design of an interactive tool inspired by the Persona method [26] to assess student motivation dynamically and within the context. While a common method to assess motivation is using questionnaires, administering them before the study asks the question out of context. On the other hand, administering questionnaires *during* an interaction can be intrusive [117]. There are some other non-intrusive approaches, such as using physiological data [76] or natural language-based approaches [117] or analyzing log files [56], but these approaches lack student agency in reporting their motivation. By considering student agency in reporting student motivation, we aim to capture the dynamic and individualized nature of motivation, taking into account students' choices and preferences. Therefore, we believe the interactive tool can serve as an alternative way to assess student motivation within the context, giving the s agency.

The primary target of the persona method is to identify representative target users from a large number of heterogeneous users and use them to guide product development. One of the critiques of the persona method is that during the persona design process, the target users are often not directly involved, potentially impacting the credibility of the personas [7]. Additionally, despite the benefit of including users in the process, involving middle school children remains a challenge due to budget constraints or difficulty in recruiting, which often leads the designers to create personas based on their own experience or input from caregivers [109]. In this thesis, we addressed this by conducting a pair of co-design sessions, including middle school students, to verify the personas developed by the researchers. This allowed the students to bring their own perspectives and experience to the process, ensuring credibility. A similar approach was made by [7] where the authors involved the graduate students in the persona design process through co-design sessions to develop a smartphone application to support their needs. Furthermore, once a set of personas are developed, they are usually shared with the design team or researchers for product development but rarely used by the target users. For example, in the 'Living Persona' approach by [118], students interacted with a single persona, but it was enacted by their teacher with the goal of improving teacher-student interaction. In this thesis, we have embedded an editable digital version of the personas in the digital textbook, Modelbook, where the students can directly modify the characteristics if applicable. Students use it to report their motivation before each digital collaborative activity giving them agency over it.

The initial and the revised interactive tool was implemented and deployed in a design and classroom studies, respectively. Observations from the design study informed revisions to the tool's implementation in the classroom study. In the classroom study, students used the tool before each digital collaborative activity and were given the opportunity to modify characteristics as desired. We discuss the major findings here. First, for the students who did not change any characteristics and remained the same with the system-assigned persona, it could mean the survey captured their motivation. Second, for the students who did change their motivations the first time, it possibly indicates that this contextually embedded, easy-to-understand persona narrative may lead the students to respond differently than surveys. Third, we also investigated further the different types of modifications they made. In the classroom study, the most changes were made to the help-giving self-concept and the least were made to conscientiousness. The changes in the factors can be explained by the literature that explains that self-confidence/concept behaviors are impacted by the context and environment surrounding the students [72]. On the other hand, conscientiousness (one of the big five personality traits) or familiarity are traits that may not change over a short time. Fourth, for students who changed the characteristics multiple times, in some cases, the changes were bi-directional, indicating a shift in their perception or experience. This suggests that the personas were probably capturing the dynamic aspects of student motivation and that students were aware of changes in their own motivation over time. Finally, we ran regression using the modified persona and initial (survey-based) persona in the classroom study data. We found that modified personas were more predictive of student participation, emphasizing that the persona tool captured student motivation while the initial persona did not.

Adaptive Collaborative Support. We developed collaborative support that incorporates both student motivation and the specific platform being utilized. Most ACLS systems tend to focus on student participation and provide collaborative support based on student actions or utterances. Often these systems overlook the importance of motivation in the learning process [82]. While some intelligent tutoring systems consider student motivation, they primarily focus on supporting individual students. In contrast, this thesis recognizes the significant role of motivation in knowledge-building and collaboration across different platforms. In traditional classrooms, teachers monitor student engagement and motivation to facilitate learning [38]. Therefore, it is crucial to consider motivation when designing an ACLS to enhance effective collaboration experiences across various platforms. Hence, in Chapter 5 and in Chapter 6, we investigated our fourth and final overarching research question *RQ4: How can we design adaptive collaborative support using student motivation and context?*

We designed and implemented collaborative support using a combination of prompts to encourage constructive participation and a badge system to reward student participation, guiding them toward the highest form of collaboration. Young students typically lack the social skills and vocabulary to discuss complex math concepts and relations [81], which makes collaboration more challenging. To address this, we incorporated specific design elements aimed at facilitating their participation. For instance, we included sentence starters to assist students in framing their responses and providing a structure for their contributions. Additionally, we implemented a badge system to encourage their active participation and provide recognition for their contributions. Many adaptive collaborative systems use conversational agents to provide feedback in the case of synchronous platforms [63, 3]. We made a deliberate decision not to include such agents in our design. This choice was based on the understanding that conversational agents could potentially disrupt the flow of conversation among students. We recognized that students might ignore agent interventions until they had finished their ongoing interaction, or students might face challenges in locating previous agent interventions within the chat thread, especially after a certain amount of time had passed [106]. Considering these factors, we opted for alternative approaches to support student collaboration without relying on conversational agents. We designed the same support interface for all three platforms so it is consistent across the platforms and minimizes the potential influence of confounding factors.

We summarize our findings related to adaptive support. First, the use of sentence starters proved beneficial for some students, facilitating their participation and enabling them to earn badges. Second, the design of the badges focused on positive feedback, reinforcing constructive behaviors such as providing help through explanations and asking clarifying questions. The results indicate a correlation between badges and constructive participation in Modelbook, suggesting that they may have contributed to increased engagement in that particular platform. Third, we observed that the effectiveness of badges varied across different contexts, and the appropriateness of specific badges depended on the context and task affordances. For instance, fewer badges were earned at Khan Academy, possibly due to the more demanding and time-consuming nature of authoring questions. Additionally, the use of sentence starters during fast-paced synchronous collaboration with the teachable agent may have impacted the flow of conversation. However, there were two major limitations in our analysis resulting from a design flaw. The first one has a predominant allocation for a certain type of support due to the rule-based support selection algorithm. The second one is the inability to compare the manipulation (survey-based support vs. persona-based support, described in Chapter 6) within the full technology condition due to a mismatch between a student's modification and the associated support factor for a particular platform. These limitations did not allow us to fully explore and evaluate the effectiveness of the adaptive support. Future research should address these limitations to gain a more comprehensive understanding of the impact and capabilities of adaptive support systems.

Learning Outcome. One of the important aspects of ACLS systems is to promote student learning. However, in our multiple studies, we observed a lack of learning outcomes among the students. Although the ultimate goal is to enhance student learning in mathematics, the focus of this work specifically concentrates on improving students' collaborative skills. Throughout this research, students took part in various activities, such as solving open-end problems in small groups using whiteboards, through discussion in Modelbook, posting or responding to questions in Khan Academy, and explaining problems to a teachable agent. The collaborative support focused on fostering constructive participation among students, encouraging activities such as elaborating, asking specific questions, and providing explanations to co-construct knowledge and remove misconceptions [120]. Several factors could have contributed to the lack of learning outcomes. For instance, students might have been less familiar with solving open-ended problems, which could have impacted their ability to fully engage with the collaborative activities. Additionally, the short duration of the study and the frequent switching between different activities might have limited the time available for students to adapt to the various types of problems and the different affordances of the platforms. As a consequence, students may have spent more time adjusting to the changing contexts rather than focusing on practicing help-giving, potentially leading to a lack of significant learning outcomes. To address this limitation and foster more effective learning experiences, future implementations of the ACLS system could consider providing students with longer, uninterrupted periods for collaborative activities, allowing them to develop their help-giving skills more thoroughly and improve their learning outcomes. It is worth noting that similar findings of limited immediate learning impact are not uncommon in the literature. For example, Olsen et al. [81] did not find differences in learning gains between students working individually or collaboratively. The lack of evidence in learning could be due to the short instruction time, which may not have been enough for differences in learning gains to emerge [81]. Also, the process that students go through for learning may be different when working on open-ended problems compared to step-based problems, so the tests did not reflect that.

Overall, the design of *UbiCoS* presented in this thesis serves as a base for a broader vision that a single adaptive collaborative system might be better than adaptive collaborative systems for each platform separately. Such a system considers the contextual variance of each platform. From a teacher's perspective, a single system offers a more comprehensive solution, for example, seamless integration of various platforms, providing students with consistent user experience and smoother transitions between collaborative activities. The students/teachers won't have to switch between multiple systems, reducing their cognitive load and improving overall usability. Moreover, a single system can dynamically adjust interventions by leveraging insights and feedback from one platform to other platforms. Finally, managing a single adaptive system is more efficient and scalable compared to maintaining separate intelligent systems for each platform. It reduces the complexity of system development, maintenance, and updates, making it easier to implement improvements and extend support to additional platforms in the future.

8.1 Limitations

Throughout this thesis, we encountered several challenges that impacted the scope and depth of our findings. The small sample size, both due to consent issues and limited participation in online study, restricted our ability to conduct robust statistical analyses and generalize the results to a larger population. However, we recognize that small data is common in educational settings, and our work contributes to understanding how to effectively work with such limited datasets to build explanatory models.

Another challenge we faced was the lack of evidence in student learning outcomes. It could be attributed to the short duration of the study, combined with the introduction of three new technologies, pedagogical approaches (modeling pedagogy), and varied collaborative activities; all of which may have contributed to cognitive overload for the students. Additionally, the assessment tests may not have adequately captured the specific skills and outcomes targeted in this thesis. Exploring more targeted and comprehensive assessment methods in future research would be valuable.

The manipulation within the full technology condition in the classroom study did not have any impact on students' help-giving, which can be attributed to a design limitation. Specifically, there was a mismatch between students' modification of characteristics and the characteristic prioritized for a particular platform, leading to inconclusive results, hence we merged the two conditions. The merging of conditions in the classroom study and the omission of analyzing face-to-face small group collaboration and online breakout room collaboration limited the depth of our findings. Further investigation into these contexts could provide valuable insights into student collaboration dynamics. However, conducting such analyses would require additional resources, such as video analysis. We propose this as one of our future works.

Despite these limitations, this work has shed light on the potential of explanatory modeling and adaptive support in the context of student collaboration. It has also highlighted areas for future research, including the need for larger sample sizes, improved assessment methods, and more comprehensive analysis of different collaborative contexts.

8.2 Future Work

Based on the work done in this thesis, there are many avenues for future work.

First, utilizing the findings in the four key areas of ACLS system development - technology, modeling, assessment, and support - researchers can further explore the broader vision of this thesis: *a single ACLS is better than ACLS for different platforms*. Several potential approaches can be pursued using this work as a building block: building three distinct ITS for the three platforms while ensuring the systems are similar and comparable. For instance, consistency in the user interface design may minimize the potential influence of confounding factors and provide a more robust basis for comparison. Another avenue of research involves conducting a comparable study that includes both a single unified ACLS and separate ACLS for different platforms. Researchers can then analyze and compare the results between the different groups by selecting participating schools, appropriately grouping students into different conditions, and collecting data on various outcomes such as learning gain, collaborative behavior, and student satisfaction. By leveraging the insights gained from this thesis and considering these factors, valuable knowledge can be gleaned to inform the design and implementation of adaptive learning systems tailored to different platforms. Such research endeavors can significantly contribute to advancing the field of adaptive collaborative learning support systems, ultimately enhancing educational experiences for students across diverse learning environments.

Second, there are opportunities for expanding the modeling approach used in this thesis. The current model focuses on explaining constructive participation, but it can be extended to develop models for other levels of participation, such as no participation or facilitative participation. Evaluating the utility of these models in real-world settings would provide valuable insights into their practical application. In the current version of the support system, a rule-based algorithm was employed to select the characteristics of each platform. A potential avenue for future research is to investigate how the explanatory model can suggest which factors to select for displaying support on each platform. By incorporating student collaboration history and domain knowledge, the model can better determine the likelihood of different participation levels and tailor the support accordingly. Alternative modeling approaches, such as Bayesian Knowledge Tracing or Hidden Markov Models, could be explored.

Third, there is room for improvement in the design and functionality of the interactive persona tool. One aspect that can be enhanced is the interface design, making it more intuitive and user-friendly for students to modify their personas. Furthermore, attention should be given to refining the underlying algorithms and mechanisms that link the modified persona to the selection and presentation of support on different platforms. By ensuring a clear and transparent relationship between persona modifications and the corresponding support, students will better understand how their actions influence the assistance they receive during collaboration.

There are numerous possibilities for further enhancing the design and algorithm of the support in adaptive collaborative learning systems. One avenue to explore is the development of automatic classification models for student utterances, enabling the system to provide more targeted and personalized feedback. By leveraging existing classification models and adapting them to the collaborative learning context of UbiCoS, it *might* be possible to identify patterns in students' utterances that reflect their motivation or engagement levels. This information can then be utilized to dynamically adjust the support provided, tailoring it to each student's specific needs. Additionally, the detection of student motivation from their utterances or the badges they receive holds the potential for further investigation. This information can then be used in conjunction with the interactive persona tool to inform the selection and presentation of support that aligns with each student's motivation profile. Furthermore, focusing on the design of the support based on platform affordances is an important area for future contributions. Each platform has its own unique characteristics and capabilities, and designing the support to leverage these affordances can enhance the overall collaborative experience.

Finally, there are several promising areas for future exploration in the field of adaptive collaborative learning systems: a) expanding the scope of this research by the integration of additional platforms to provide insights into the transferability and effectiveness of the model across different contexts such as online asynchronous discussion forums, or collaborative coding platforms. b) building a dashboard for teachers that provides real-time information and insights about students' collaborative activities can help to better facilitate and support students' learning. The dashboard can provide data on student participation, motivation levels, and even recommendations for intervention or support strategies based on the model's output. c) The COVID-19 pandemic has highlighted the importance of addressing access and equity issues in education. Future research should explore ways to mitigate the difficulties that arise when students participate in different environments with varying resources.

8.3 Conclusion

This chapter provides a comprehensive summary of the thesis, highlighting its contributions and implications for various fields of study. This work contributed to multiple fields of study, including Human-Computer Interaction, Computer Science, and Learning Science. One of the major contributions of this thesis is the design of a collaborative learning environment encompassing multiple platforms instead of a self-contained environment. Another key contribution is the adaptation of traditional Intelligent Tutoring Systems (ITS) to model student collaboration using individual and platform characteristics. Another significant contribution is the design of an interactive tool for the dynamic assessment of student motivation. Furthermore, the provision of personalized support across platforms, leveraging student motivation to facilitate effective collaboration, has been a prime focus of this work. We also acknowledged the limitations and described future work. By addressing these, researchers can further advance the field of adaptive collaborative learning and contribute to the development of more inclusive, effective, and supportive educational technologies.

Bibliography

[1] Exploring Metaculus's AI Track Record. `https://www.metaculus.com/notebooks/16708/exploring-metaculuss-ai-track-record`. Accessed: 2023-05-30.

[2] Samuel Abramovich, Christian Schunn, and Ross Mitsuo Higashi. Are badges useful in education?: It depends upon the type of badge and expertise of learner. *Educational Technology Research and Development*, 61:217–232, 2013.

[3] David Adamson, Gregory Dyke, Hyeju Jang, and Carolyn Penstein Rosé. Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, 24(1):92–124, 2014.

[4] Lenore Ellen Adie, Jill Willis, and Fabienne Michelle Van der Kleij. Diverse perspectives on student agency in classroom assessment. *The Australian Educational Researcher*, 45(1):1–12, 2018.

[5] Ishrat Ahmed, Areej Mawasi, Shang Wang, Ruth Wylie, Yoav Bergner, Amanda Whitehurst, and Erin Walker. Investigating help-giving behavior in a cross-platform learning environment. In *International Conference on Artificial Intelligence in Education*, pages 14–25. Springer, 2019.

[6] Hua Ai, Marietta Sionti, Yi-Chia Wang, and Carolyn Rose. Finding transactive contributions in whole group classroom discussions. 2010.

[7] Charlotte Albrechtsen, Majbrit Pedersen, Nicholai Friis Pedersen, and Tine Wirenfeldt Jensen. Proposing co-design of personas as a method to heighten validity and engage users: a case from higher education. *International Journal of Sociotechnology and Knowledge Development (IJSKD)*, 8(4):55–67, 2016.

[8] Fatima Ali Amer Jid Almahri, David Bell, and Mahir Arzoky. Personas design for conversational systems in education. In *Informatics*, volume 6, page 46. Multidisciplinary Digital Publishing Institute, 2019.

[9] Alissa N Antle. Child-based personas: need, ability and experience. *Cognition, Technology & Work*, 10(2):155–166, 2008.

[10] Margarita Azmitia and Ryan Montgomery. Friendship, transactive dialogues, and the development of scientific reasoning. *Social development*, 2(3):202–221, 1993.

[11] Nilufar Baghaei, Antonija Mitrovic, and Warwick Irwin. Supporting collaborative learning and problem-solving in a constraint-based cscl environment for uml class diagrams. *International Journal of Computer-Supported Collaborative Learning*, 2(2):159–190, 2007.

[12] Claudio Barbaranelli, Gian Vittorio Caprara, Annarita Rabasca, and Concetta Pastorelli. A questionnaire for measuring the big five in late childhood. *Personality and individual differences*, 34(4):645–664, 2003.

[13] Jean-Louis Berger and Stuart A Karabenick. Motivation and students' use of learning strategies: Evidence of unidirectional effects in mathematics classrooms. *Learning and instruction*, 21(3):416–428, 2011.

[14] Marvin W Berkowitz and John C Gibbs. Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly (1982-)*, pages 399–410, 1983.

[15] Przemyslaw Biecek and Tomasz Burzykowski. *Explanatory model analysis: explore, explain, and examine predictive models*. CRC Press, 2021.

[16] Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001.

[17] Glenn W Brier et al. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

[18] Patrick Buckley and Elaine Doyle. Gamification and student motivation. *Interactive learning environments*, 24(6):1162–1175, 2016.

[19] John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3):294–320, 2013.

[20] Cheng-Huan Chen and Chiung-Hui Chiu. Collaboration scripts for enhancing metacognitive self-regulation and mathematics literacy. *International Journal of Science and Mathematics Education*, 14(2):263–280, 2016.

[21] Wing Sum Cheung, Khe Foon Hew, and Connie Siew Ling Ng. Toward an understanding of why students contribute in asynchronous online discussions. *Journal of Educational Computing Research*, 38(1):29–50, 2008.

[22] Michelene TH Chi. Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, 1(1):73–105, 2009.

[23] Namok Choi. Self-efficacy and self-concept as predictors of college students' academic performance. *Psychology in the Schools*, 42(2):197–205, 2005.

[24] Chientzu Candace Chou and Shi-Jen He. The effectiveness of digital badges on student online contributions. *Journal of Educational Computing Research*, 54(8):1092–1116, 2017.

[25] Elizabeth G Cohen. Restructuring the classroom: Conditions for productive small groups. *Review of educational research*, 64(1):1–35, 1994.

[26] Alan Cooper. The inmates are running the asylum. In *Software-Ergonomie'99*, pages 17–17. Springer, 1999.

[27] PL Cornelius and MS Seyedsadr. Estimation of general linear-bilinear models for two-way tables. *Journal of Statistical Computation and Simulation*, 58(4):287–322, 1997.

[28] Gayle V Davidson-Shivers, Lin Y Muilenburg, and Erica J Tanner. How do students participate in synchronous and asynchronous online discussions? *Journal of Educational Computing Research*, 25(4):351–366, 2001.

[29] Maria de los Angeles Constantino-Gonzalez, Daniel D Suthers, and José G Escamilla de los Santos. Coaching web-based collaborative learning based on problem solution differences and participation. *International Journal of Artificial Intelligence in Education*, 13(2-4):263–299, 2003.

[30] Angel De Vicente. Towards tutoring systems that detect students' motivation: an investigation. 2003.

[31] Angel de Vicente and Helen Pain. Motivation self-report in its.

[32] Angel De Vicente and Helen Pain. Motivation diagnosis in intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*, pages 86–95. Springer, 1998.

[33] Matt Dennis, Judith Masthoff, and Chris Mellish. Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education*, 26(3):877–931, 2016.

[34] Dejana Diziol, Erin Walker, Nikol Rummel, and Kenneth R Koedinger. Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, 22:89–102, 2010.

[35] Zoltán Dörnyei. Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British journal of educational psychology*, 70(4):519–538, 2000.

[36] Toby Dragon and Beverly Park Woolf. Guidance and collaboration strategies in ill-defined domains.

[37] Benedict du Boulay. Intelligent tutoring systems that adapt to learner motivation. *Tutoring and intelligent tutoring systems*, pages 103–128, 2018.

[38] Benedict du Boulay and Teresa Del Soldato. Implementation of motivational tactics in tutoring systems: 20 years on. *International Journal of Artificial Intelligence in Education*, 26(1):170–182, 2016.

[39] Teresa Garcia Duncan and Wilbert J McKeachie. The making of the motivated strategies for learning questionnaire. *Educational psychologist*, 40(2):117–128, 2005.

[40] Julian Elliott*. Multimethod approaches in educational research. *International Journal of Disability, Development and Education*, 51(2):135–149, 2004.

[41] Usef Faghihi, Albert Brautigam, Kris Jorgenson, David Martin, Angela Brown, Elizabeth Measures, and Sioui Maldonado-Bouchard. How gamification applies for educational purpose specially with college algebra. *Procedia Computer Science*, 41:182–187, 2014.

[42] John W Fantuzzo, Judith A King, and Lauren R Heller. Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, 84(3):331, 1992.

[43] Lynn S Fuchs, Douglas Fuchs, Carol L Hamlett, Norris B Phillips, Kathy Karns, and Suzanne Dutka. Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 97(3):223–249, 1997.

[44] Sara M Fulmer and Jan C Frijters. A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review*, 21(3):219–246, 2009.

[45] Xun Ge and Susan M Land. Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational technology research and development*, 51(1):21–38, 2003.

[46] Gahgene Gweon, Carolyn Rose, Regan Carey, and Zachary Zaiss. Providing support for adaptive scripting in an on-line collaborative learning environment. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 251–260, 2006.

[47] Lassi Haaranen, Petri Ihantola, Lasse Hakulinen, and Ari Korhonen. How (not) to introduce badges to online exercises. In *Proceedings of the 45th ACM technical symposium on Computer science education*, pages 33–38, 2014.

[48] Ijaz Ul Haq, Aamir Anwar, Ikram Ur Rehman, Waqar Asif, Drishty Sobnath, Hafiz Husnain Raza Sherazi, and Moustafa M Nasralla. Dynamic group formation with intelligent tutor collaborative learning: a novel approach for next generation collaboration. *IEEE Access*, 9:143406–143422, 2021.

[49] Andreas Harrer, Bruce M McLaren, Erin Walker, Lars Bollen, and Jonathan Sewall. Creating cognitive tutors for collaborative learning: steps toward realization. *User Modeling and User-Adapted Interaction*, 16:175–209, 2006.

[50] Rachel Harsley, Barbara Di Eugenio, Nick Green, Davide Fossati, and Sabita Acharya. Integrating support for collaboration in a computer science intelligent tutoring system. In *Intelligent Tutoring Systems: 13th International Conference, ITS 2016, Zagreb, Croatia, June 7-10, 2016. Proceedings 13*, pages 227–233. Springer, 2016.

[51] Khe Foon Hew, Wing Sum Cheung, and Connie Siew Ling Ng. Student contribution in asynchronous online discussion: A review of the research and empirical exploration. *Instructional science*, 38(6):571–606, 2010.

[52] Cindy E Hmelo-Silver and Howard S Barrows. Facilitating collaborative knowledge building. *Cognition and instruction*, 26(1):48–94, 2008.

[53] Judith Israel and Robert Aiken. Supporting collaborative learning with an intelligent web-based system. *International Journal of Artificial Intelligence in Education*, 17(1):3–40, 2007.

[54] Jane Jackson, Larry Dukerich, and David Hestenes. Modeling instruction: An effective model for science education. *Science Educator*, 17(1):10–17, 2008.

[55] H Jeong. Knowledge co-construction during collaborative learning. 1999.

[56] Jeffrey Johns and Beverly Woolf. A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the national conference on artificial intelligence*, volume 21, page 163. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[57] David W Johnson and Roger T Johnson. *Learning together and alone: Cooperative, competitive, and individualistic learning.* Prentice-Hall, Inc, 1987.

[58] David W Johnson and Roger T Johnson. Cooperative learning and achievement. 1990.

[59] Genevieve Marie Johnson. Synchronous and asynchronous text-based cmc in educational contexts: A review of recent research. *TechTrends*, 50(4):46–53, 2006.

[60] Mahesh Joshi and Carolyn Penstein Rosé. Using transactivity in conversation for summarization of educational dialogue. In *Workshop on Speech and Language Technology in Education*, 2007.

[61] Anastasios Karakostas and S Demetriadis. Enhancing collaborative learning through dynamic forms of support: the impact of an adaptive domain-specific support strategy. *Journal of Computer Assisted Learning*, 27(3):243–258, 2011.

[62] Alison King. Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American educational research journal*, 31(2):338–368, 1994.

[63] Rohit Kumar, Carolyn Penstein Rosé, Yi-Chia Wang, Mahesh Joshi, and Allen Robinson. Tutorial dialogue as adaptive collaborative learning support. *Frontiers in artificial intelligence and applications*, 158:383, 2007.

[64] Annette M La Greca, Susan Kraslow Dandes, Patricia Wick, Kimberly Shaw, and Wendy L Stone. Development of the social anxiety scale for children: Reliability and concurrent validity. *Journal of Clinical Child Psychology*, 17(1):84–91, 1988.

[65] Ard W Lazonder, Pascal Wilhelm, and Susanne AW Ootes. Using sentence openers to foster student interaction in computer-mediated learning environments. *Computers & Education*, 41(3):291–308, 2003.

[66] Joey J Lee and Jessica Hammer. Gamification in education: What, how, why bother? *Academic exchange quarterly*, 15(2):146, 2011.

[67] Lasse Lipponen. Exploring foundations for computer-supported collaborative learning. 2002.

[68] Chen-Chung Liu and Chin-Chung Tsai. An analysis of peer interaction patterns as discoursed by on-line small group problem-solving activity. *Computers & Education*, 50(3):627–639, 2008.

[69] Nichola Lubold, Heather Pon-Barry, and Erin Walker. Naturalness and rapport in a pitch adaptive learning companion. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 103–110. IEEE, 2015.

[70] Heng Luo, Tingting Yang, Jin Xue, and Mingzhang Zuo. Impact of student agency on learning performance and learning experience in a flipped classroom. *British Journal of Educational Technology*, 50(2):819–831, 2019.

[71] Herbert W Marsh. *Self-description Questionnaire III: SDQ III.* Self-concept Enhancement and Learning Facilitation (SELF) Research Centre . . . , 1999.

[72] Herbert W Marsh, Ulrich Trautwein, Oliver Lüdtke, Olaf Köller, and Jürgen Baumert. Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of personality*, 74(2):403–456, 2006.

[73] Areej Mawasi, Ishrat Ahmed, Erin Walker, Shang Wang, Zeynep Marasli, Amanda Whitehurst, and Ruth Wylie. Using design-based research to improve peer help-giving in a middle school math classroom. 2020.

[74] Robert R McCrae and Paul T Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81, 1987.

[75] Bruce M McLaren, Oliver Scheuer, and Jan Mikšátko. Supporting collaborative learning and e-discussions using artificial intelligence techniques. *International Journal of Artificial Intelligence in Education*, 20(1):1–46, 2010.

[76] Scott W McQuiggan and James C Lester. Diagnosing self-efficacy in intelligent tutoring systems: an empirical study. In *International Conference on Intelligent Tutoring Systems*, pages 565–574. Springer, 2006.

[77] Antonija Mitrovic, Stellan Ohlsson, and Devon K Barrow. The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education*, 60(1):264–272, 2013.

[78] Maryam Naghizadeh and Hadi Moradi. A model for motivation assessment in intelligent tutoring systems. In *2015 7th Conference on Information and Knowledge Technology (IKT)*, pages 1–6. IEEE, 2015.

[79] Amalya Nattiv. Helping behaviors and math achievement gain of students using cooperative learning. *The Elementary School Journal*, 94(3):285–297, 1994.

[80] Amy Ogan, Samantha Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. Rudeness and rapport: Insults and learning gains in peer tutoring. In *International Conference on Intelligent Tutoring Systems*, pages 11–21. Springer, 2012.

[81] Jennifer K Olsen, Nikol Rummel, and Vincent Aleven. Investigating effects of embedding collaboration in an intelligent tutoring system for elementary school students. Singapore: International Society of the Learning Sciences, 2016.

[82] Fidelia A Orji and Julita Vassileva. Modelling and quantifying learner motivation for adaptive systems: Current insight and future perspectives. In *International Conference on Human-Computer Interaction*, pages 79–92. Springer, 2021.

[83] Alejandro Ortega-Arranz, Erkan Er, Alejandra Martínez-Monés, Miguel L Bote-Lorenzo, Juan I Asensio-Pérez, and Juan A Muñoz-Cristóbal. Understanding student behavior and perceptions toward earning badges in a gamified mooc. *Universal Access in the Information Society*, 18(3):533–549, 2019.

[84] Murat Oztok, Daniel Zingaro, Clare Brett, and Jim Hewitt. Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, 60(1):87–94, 2013.

[85] A Sullivan Palincsar. Social constructivist perspectives on teaching and learning. *Annual review of psychology*, 49(1):345–375, 1998.

[86] Aannemarie Sullivan Palinscar and Ann L Brown. Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction*, 1(2):117–175, 1984.

[87] Alexandros Paramythis. Adaptive support for collaborative learning with ims learning design: are we there yet. In *Proceedings of the Workshop on Adaptive Collaboration Support, held in conjunction with the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Hannover, Germany*, pages 17–29. Citeseer, 2008.

[88] Philip David Parker, Herbert W Marsh, Joseph Ciarrochi, Sarah Marshall, and Adel Salah Abduljabbar. Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, 34(1):29–48, 2014.

[89] Paul R Pintrich and Elisabeth V De Groot. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 82(1):33, 1990.

[90] Stephen R Porter, Michael E Whitcomb, and William H Weitzer. Multiple surveys of students and survey fatigue. *New directions for institutional research*, 2004(121):63–73, 2004.

[91] Lei Qu, Ning Wang, and W Lewis Johnson. Using learner focus of attention to detect learner motivation factors. In *User Modeling 2005: 10th International Conference, UM 2005, Edinburgh, Scotland, UK, July 24-29, 2005. Proceedings 10*, pages 70–73. Springer, 2005.

[92] Marta C Rosatelli and John A Self. A collaborative case study system for distance learning. *International Journal of Artificial Intelligence in Education*, 14(1):97–125, 2004.

[93] Rod D Roscoe and Michelene TH Chi. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of educational research*, 77(4):534–574, 2007.

[94] Carolyn Rosé, Yi-Chia Wang, Yue Cui, Jaime Arguello, Karsten Stegmann, Armin Weinberger, and Frank Fischer. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International journal of computer-supported collaborative learning*, 3(3):237–271, 2008.

[95] Carolyn P Rosé, Elizabeth A McLaughlin, Ran Liu, and Kenneth R Koedinger. Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, 50(6):2943–2958, 2019.

[96] Borhan Samei, Haiying Li, Fazel Keshtkar, Vasile Rus, and Arthur C Graesser. Context-based speech act classification in intelligent tutoring systems. In *International conference on intelligent tutoring systems*, pages 236–241. Springer, 2014.

[97] Mangalam Sankupellay, Erica Mealy, Christoph Niesel, and Richard Medland. Building personas of students accessing a peer-facilitated support for learning program. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, pages 412–416, 2015.

[98] Pedro Bispo Santos, Caroline Verena Bhowmik, and Iryna Gurevych. Avoiding bias in students' intrinsic motivation detection. In *Intelligent Tutoring Systems: 16th International Conference, ITS 2020, Athens, Greece, June 8–12, 2020, Proceedings*, pages 89–94. Springer, 2020.

[99] Daniel L Schwartz and Taylor Martin. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and instruction*, 22(2):129–184, 2004.

[100] Galit Shmueli. To explain or to predict? 2010.

[101] Robert E Slavin. *Cooperative Learning. Research on Teaching Monograph Series.* ERIC, 1983.

[102] Robert E Slavin. Team-assisted individualization. In *Learning to cooperate, cooperating to learn*, pages 177–209. Springer, 1985.

[103] Faiza Tahir, Antonija Mitrovic, and Valerie Sotardi. Investigating the causal relationships between badges and learning outcomes in sql-tutor. *Research and Practice in Technology Enhanced Learning*, 17(1):1–23, 2022.

[104] Martha Tapia and George E Marsh II. An instrument to measure mathematics attitudes. *Academic exchange quarterly*, 8(2):16–21, 2004.

[105] Stephanie D Teasley. Talking about reasoning: How important is the peer in peer collaboration? In *Discourse, tools and reasoning*, pages 361–384. Springer, 1997.

[106] Stergios Tegos, Stavros Demetriadis, and Thrasyvoulos Tsiatsos. A configurable conversational agent to trigger students' productive dialogue: a pilot study in the call domain. *International Journal of Artificial Intelligence in Education*, 24:62–91, 2014.

[107] David R Thomas. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, 27(2):237–246, 2006.

[108] Jennifer Tsan. *Developing Adaptive Support for Collaborative Problem Solving Between Upper-Elementary Computer Science Learners*. North Carolina State University, 2020.

[109] Shirley Varela, Candida Hall, and Hee Jin Bang. Creating middle school child-based personas for a digital math practice application. In *EdMedia+ Innovate Learning*, pages 532–537. Association for the Advancement of Computing in Education (AACE), 2015.

[110] Aurora Vizcaíno. A simulated student can improve collaborative learning. *International Journal of Artificial Intelligence in Education*, 15(1):3–40, 2005.

[111] Aurora Vizcaíno, Juan Contreras, Jesús Favela, and Manuel Prieto. An adaptive, collaborative environment to develop good habits in programming. In *Intelligent Tutoring Systems*, volume 1839, pages 262–271. Springer, 2000.

[112] James F Voss. On the solving oi iii-structured problems. *The nature of expertise*, 261, 2014.

[113] Lev Semenovich Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[114] Erin Walker, Amy Ogan, Vincent Aleven, and Chris Jones. Two approaches for providing adaptive support for discussion in an ill-defined domain. *Intelligent Tutoring Systems for Ill-Defined Domains: Assessment and Feedback in Ill-Defined Domains*, 1, 2008.

[115] Erin Walker, Nikol Rummel, and Kenneth R Koedinger. Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, 24(1):33–61, 2014.

[116] Feng Wang and Michael J Hannafin. Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4):5–23, 2005.

[117] Arthur Ward, Diane Litman, and Maxine Eskenazi. Predicting change in student motivation by measuring cohesion between tutor and student. In *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 136–141, 2011.

[118] Bruno Warin, Christophe Kolski, and Claudine Toffolon. Living persona technique applied to hci education. In *2018 IEEE Global Engineering Education Conference (EDUCON)*, pages 51–59. IEEE, 2018.

[119] Noreen M Webb. Peer interaction and learning in small groups. *International journal of Educational research*, 13(1):21–39, 1989.

[120] Noreen M Webb and Sydney Farivar. Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal*, 31(2):369–395, 1994.

[121] Noreen M Webb and Ann Mastergeorge. Promoting effective helping behavior in peer-directed groups. *International journal of educational research*, 39(1-2):73–97, 2003.

[122] Allan Wigfield. Expectancy-value theory of achievement motivation: A developmental perspective. *Educational psychology review*, 6(1):49–78, 1994.

[123] Dezhi Wu and Starr Roxanne Hiltz. Predicting learning from asynchronous online discussions. *Journal of Asynchronous Learning Networks*, 8(2):139–152, 2004.

[124] Kui Xie and Amy C Bradshaw. Using question prompts to support ill-structured problem solving in online peer collaborations. *International Journal of Technology in Teaching and Learning*, 4(2):148–165, 2008.