Using Design-Based Research to Improve Peer Help-Giving in a Middle School Math Classroom

Areej Mawasi, Arizona State University, amwassi@asu.edu
Ishrat Ahmed, University of Pittsburgh, ISA14@pitt.edu
Erin Walker, University of Pittsburgh, EAWALKER@pitt.edu
Shang Wang, Arizona State University, swang158@asu.edu
Zeynep Marasli, University of Pittsburgh, beyza.marasli@gmail.com
Amanda Whitehurst, STEMteachersPHX, albinomosquito116@gmail.com
Ruth Wylie, Arizona State University, ruth.wylie@asu.edu

Abstract: Computer-Supported Collaborative Learning (CSCL) environments are often designed to support collaboration within a single digital platform. However, with the growth of technology in classrooms, students often find themselves working in multiple contexts (i.e., a student might work face-to-face with a peer on one task and then move to engaging in an online discussion for homework). We have created a CSCL environment that aims to support student help-giving across a variety of digital platforms. This paper describes three cycles of a design-based research study that aims to design a system to support help-giving and improve interaction quantity and quality across different contexts as well as to better understand whether students benefit by the addition of multiple contexts. The paper shares major refinements across the three cycles that worked to balance research, pedagogical, and technological goals to improve students' help-giving behavior in a middle-school mathematics classroom.

Introduction

When students engage in collaborative learning, they have the opportunity to build on each other's knowledge and develop new ideas, improving domain and metacognitive learning outcomes (Chi & Wylie, 2014). However, many students do not possess the necessary skills to benefit from collaboration. Students must be able to elaborate on their knowledge and ask specific questions, and often they need to feel an affinity and rapport towards their group mates (Webb & Farivar, 1994; Graves & Klawe, 1997). Research in math education suggests that while engaging learners in help-giving behaviors such as explanation can lead to positive learning outcomes, not all students benefit from such activities, because factors such as prior knowledge, competence, or identities may influence their behavior (e.g., Esmonde, 2009; Webb & Farivar, 1994). In addition, peer interaction requires classroom management strategies from teachers to orchestrate their classrooms, create norms for collaborative discourse, and balance learners' needs, curricular needs, and pedagogy (e.g., Wentzel & Watkins, 2011; Dillenbourg, 2013; Roschelle et al., 2013). Finally, as learning expands across multiple activity systems (e.g., Engeström & Sannino, 2010), understanding how distributed collaborative learning activities occur is essential for designing systems to mediate these collaborations.

Computer-Supported Collaborative Learning (CSCL) environments can be used to address these challenges. They can open the door for innovations that bridge pedagogy and educational technology and they can allow multiple ways to examine learning activities across multiple settings. These environments provide a variety of affordances for enabling students to communicate with collaborators, facilitating the exchange of resources both in and out of the classroom, and allowing students to engage in knowledge co-construction, monitoring, and regulation of activities (Jeong & Hmelo-Silver, 2016; Scardamalia & Bereiter, 1994; Drie, Boxtel, & Erkens, 2005). However, existing platforms are often constrained to a specific context (e.g., students interacting with each other via chat or face-to-face). With the increasing presence of collaboration technologies in formal learning environments, it is important to understand how students collaborate across a variety of technologies and contexts. For example, do students behave the same when helping peers via chat versus helping strangers on an online forum? These insights can help support learners' cross-platform interactions (Ahmed et al., 2019).

Our work builds on research on distributed learning to create a suite of CSCL environments we term UbiCoS (Ubiquitous Collaboration Support). UbiCoS aims to improve students' help-giving behaviors across multiple activities and platforms. We focus on students helping each other in a middle school mathematics class where students have opportunities to answer questions posed by their peers and provide feedback on their thinking. Help-giving is a promotive interaction for collaboration, has cognitive benefits as students engage in co-constructing knowledge, and has motivational benefits as providing help can increase self-efficacy (e.g., Webb & Farivar, 1994),. 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. UbiCoS was designed and implemented by an interdisciplinary team including computer scientists, learning scientists, middle school students, and middle school teachers (Ahmed et al., 2019).

In this paper, we describe a system that integrates technology and curriculum to support help-giving and improve interaction quantity and quality across digital contexts. In addition, we describe how these contexts can differentially serve students' help-giving needs. We apply a design-based research methodology involving three major design cycles in a middle school math classroom (Barab & Squire, 2004; Wang & Hannafin, 2005; Sandoval, 2014). Each cycle consisted of technology development and curriculum refinement, followed by a week of classroom implementation. This research seeks to understand and support collaborative learning across multiple spaces, balancing research goals and the pedagogical needs of the classroom. It contributes to collaboration theory by investigating how students interact across contexts and provides practical tips for implementing these technologies.

Initial design of UbiCoS

To orchestrate help-giving activities across multiple settings, we pursued two major directions: (1) designing curriculum to promote productive interactions and relationship-building in our face-to-face and digital environments, and (2) designing technology to reify the curriculum and support these activities. To accomplish these goals, our team includes a co-designer teacher who had taught middle school for 13 years and was trained in modeling pedagogy. Modeling pedagogy is based on constructivism, where students engage in small-group, open-ended investigation of a concrete problem that provides the basis for the development of a conceptual model. Students express their models on whiteboards, give feedback on other groups' whiteboards, and ultimately have a whole-class discussion to arrive on a set of principles related to the models (Jackson, Dukerich, & Hestenes, 2008). 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 (Hmelo-Silver & Barrows, 2008) and invention as preparation for future learning (Schwartz & Martin, 2004). We chose modeling curriculum because it encourages collaborative interactions, develops learners' sense of community, and engages students with math practices such as explanation and critique of 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 that aligned with Common Core and state standards for eighth-grade mathematics: ratios and proportions, volume and surface area, and linear functions.

To complement the curriculum and the small-group and whole-class discussions, we introduced three digital contexts in which students could interact. The first was a digital textbook called ModelBook that we developed to contain the curricular materials (e.g., question prompts, homework assignments), allow students to log their work (e.g., students could upload photos of their whiteboards), and enable students to interact digitally with their classmates. 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 upload work they completed in face-to-face groups and evaluate, critique, and provide feedback to others through discussion (see Figure 1). Another tool for whole-class discussion is called the 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 the photo of their whiteboard to a digital gallery, and then engage in a digital discussion about their whiteboard with another group of students.

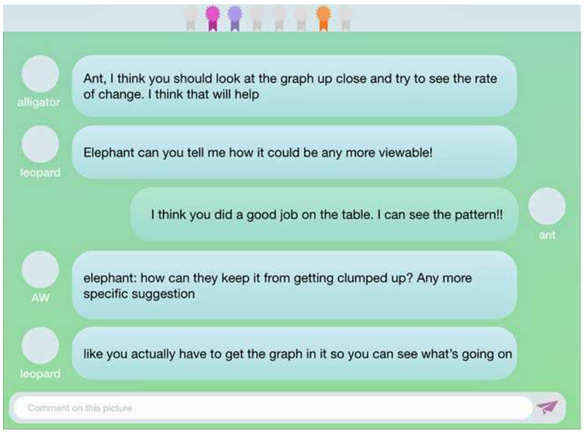

Figure 1. ModelBook Gallery Discussion.

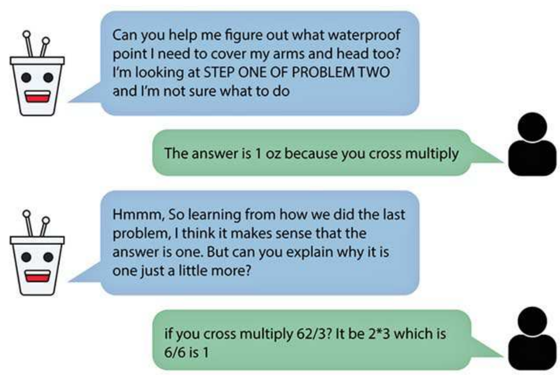

Figure 2. Sample teachable agent interaction.

The second digital context was Khan Academy (https://www.khanacademy.org), an asynchronous public question-answer platform where students watch short videos and participate in Q&A forums. By answering questions, students articulate their understanding and engage in help-giving behavior with the broader public. The third digital context was a Virtual Teachable Agent, a desktop version adapted from Lubold, et al. (2019). In this system, students work individually with an agent to help it solve step-based mathematics problems. Students follow a worked example and explain each step to the agent. The agent responds in spoken dialogue, and if the student provides the correct answer but not an explanation, the agent prompts the students to elaborate (see Figure 2). Through these interactions, students practice help-giving skills in a context where domain knowledge is less of a factor (because of the provided worked examples) and social anxiety is lessened since students are interacting with an agent, rather than a peer.

Each of the four contexts (three digital contexts plus 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 in a face-to-face discussion or ModelBook activity, and less likely to answer a stranger's question on Khan Academy. A student with low math ability who fears making a mistake in front of a classmate may contribute less when interacting with a peer and contribute more often on Khan Academy or with the Teachable Agent. Contexts also varied by interaction modality (speech versus text) and timing (synchronous and asynchronous). For this paper, we focus on digital interactions.

Methods

Participants and procedures

This research was conducted in a middle school in the Southwestern United States. The three cycles took place in an eighth-grade math classroom with the same group of students and were part of the regular curriculum. The number of students who were present and consented to participate differed from one cycle to another (Cycle 1: 20, Cycle 2: 26, Cycle 3: 24). Out of the 26 total participants, 15 were male, 9 were female, and 2 did not report. Self-reported ethnicity was: Hispanic (n=8), White (n=8), African American (n=1), Native American (n=2), Other (n=5), and Unknown (n=2).

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. 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 contexts described above, students participated in a variety of activities, including whole-class discussion, direct instruction, individual work within ModelBook, and hands-on activities (described in more detail below).

Measures

For each cycle, we assessed learning outcomes using two isomorphic forms for the pretest and posttest. We counterbalanced the forms (i.e., half the students received Form A for the pretest and half received Form B), and students received the opposite form for the posttest. 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. As part of the broader study, we collected data on students' small-group face-to-face interactions and qualitative perceptions of using the system; however, within the scope of this paper, we focus on students' digital interactions. Specifically, we collected interaction data to address how students were interacting across contexts and to determine if student interactions changed across cycles. For each cycle, we measured quantity and quality of interactions within Khan Academy, ModelBook, and the Teachable Agent. Using both quantity and quality of interactions allowed us to look at moments where learners were engaged with the system and to examine the quality of their contributions. We developed an ordered coding scheme inspired by elaborated help (Webb & Farivar, 1994) and transactive reasoning (Berkowitz & Gibbs, 1983):

- 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. ModelBook example: "I agree". Khan Academy example: "I have no clue why."

- 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!"). ModelBook example: "I like the graph and the notes you added." Khan Academy example: "yes you are correct because he was correct"
- 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 explanation, or asking a specific clarification question. ModelBook example: "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" Khan Academy example: "a constant function is basically when its always producing the same number as the outcome for most of the time"

For ModelBook and Khan Academy, the unit of analysis was a student's post. We coded the highest category achieved within a given problem (e.g., if a student had a single constructive utterance, their code was constructive participation). To compute reliability of codes, two raters independently coded 40% of the ModelBook data, with Intra-Class Correlation Coefficients (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.

Design-based research cycles

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. These digital discussions were often followed by whole class face-to-face "board meetings." 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 understanding of ratios and proportions to model the speed of a moving car. 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. On three out of the five days, students were assigned to watch videos on Khan Academy and participate in online discussions 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 in order to make constructive posts, such as disagreeing with another student's post and explaining why. 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.

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 be 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. 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 1, 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 1: Student 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, 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

The purpose of this cycle was to implement changes based on Cycle 1 in order to improve learning gains, improve interaction quality, and further explore differential participation patterns across the different contexts. 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 to prompt higher-quality conversation.* Instead of having students engage in several different Gallery discussions, we put them in a single discussion group and had them discuss a few images with a set group of students.
4. *Implemented badges to prompt higher-quality conversation*. We began awarding badges when the system detected high-quality posts in the ModelBook and Khan Academy contexts. 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 wasn’t 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 volume of the different shapes. At the end of the week, the hands-on activity asked students to calculate how many conical and spherical 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.

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 and a mean of 2.09 posts per activity for the Chats. In Khan Academy, students made a mean of 0.80 posts per assignment. We should note that some students posted directly on Khan Academy while others posted 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, 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 on 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 in the activity at all. For the Teachable Agent, students completed roughly five out of the six problems.

Table 2: Student 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%

With respect to interaction quality, we also see a decline in Cycle 2 compared to Cycle 1 with respect to ModelBook (See Table 2). Note that there was one Chat activity where only 13 students participated, and there was a lot of off-topic (minimal participation) conversation. 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

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. Therefore, 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 assignment.

In Cycle 3, students continued to work in their face-to-face groups and have digital discussions using ModelBook. Students began with the hands-on 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.

Table 3: Student interactions 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%	34.4%
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%

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 on 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). 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 3). 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.

Discussion and implications

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 and 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 helped us to develop 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 time 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, 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 influenced 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 in ModelBook but lower than average on Khan Academy, and vice versa. This suggests that by including multiple contexts, we are serving diverse learners. 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.

References

Ahmed, I., Mawasi, A., Wang, S., Wylie, R., Bergner, Y., Whitehurst, A., & Walker, E. (2019, June). Investigating help-giving behavior in a cross-platform learning environment. In *International Conference on Artificial Intelligence in Education* (pp. 14-25). Springer, Cham.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1-14.

Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. Merrill-Palmer Quarterly (1982-), 399-410.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485-492.

Drie, J. V., Boxtel, C. V., Erkens, G., & Kanselaar, G. (2005). Using representational tools to support historical reasoning in computer-supported collaborative learning. *Technology, Pedagogy and Education*, 14(1), 25-41

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational research review*, *5*(1), 1-24.

Esmonde, I. (2009). Mathematics learning in groups: Analyzing equity in two cooperative activity structures. *The Journal of the Learning Sciences*, 18(2), 247-284.

Graves, D., & Klawe, M. (1997, December). Supporting learners in a remote CSCL environment: The importance of task and communication. In *Proceedings of the 2nd International Conference on Computer Support for Collaborative Learning* (pp. 65-75). International Society of the Learning Sciences.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and instruction*, 26(1), 48-94.

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: An effective model for science education. *Science Educator*, 17(1), 10-17.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, 51(2), 247-265.

Lubold, N., Walker, E., Pon-Barry, H., & Ogan, A. (2019, June). Comfort with Robots Influences Rapport with a Social, Entraining Teachable Robot. In *International Conference on Artificial Intelligence in Education* (pp. 231-243). Springer, Cham.

Roschelle, J., Dimitriadis, Y., & Hoppe, U. (2013). Classroom orchestration: synthesis. *Computers & Education*, 69, 523-526.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences*, *23*(1), 18-36.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the learning sciences*, 3(3), 265-283.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129-184.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4), 5-23.

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal*, 31(2), 369-395.

Wentzel, K. R., & Watkins, D. E. (2011). Instruction based on peer interactions. *Handbook of research on learning and instruction*, 322-343.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 1736103.