Exploring Human–AI Control over Dynamic Transitions between Individual and Collaborative Learning

Vanessa Echeverria[1[0000-0002-2022-9588]], Kenneth Holstein[1], Jennifer Huang[1], Jonathan Sewall[1], Nikol Rummel[2] and Vincent Aleven[1]

[1] Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA
[2] Institute of Educational Research, Ruhr-University Bochum, Bochum, Germany
{vie,kjholste,jjhuang1,sewall,aleven}@andrew.cmu.edu
nikol.rummel@rub.de

Abstract. *Dynamically transitioning* between individual and collaborative learning activities during a class session (i.e., in an un-planned way, as-the-need-arises) may have advantages for students. Existing orchestration tools are not designed to support such transitions. This work reports findings from a technology probe study that explored alternative designs for classroom co-orchestration support for dynamically transitioning between individual and collaborative learning, focused on how control over the transitions should be divided or shared among teachers, students, and orchestration system. This study involved 1) a pilot in an authentic classroom scenario with AI support for individual and collaborative learning; and 2) design workshops and interviews with students and teachers. Findings from the study suggest the need for *hybrid control between students, teachers, and AI systems* over transitions as well as for adaptivity and/or adaptability for different classrooms, teachers, and students' prior knowledge. This study is the first to explore human–AI control over dynamic transitions between individual and collaborative learning in actual classrooms.

Keywords: individual and collaborative learning, human-AI co-orchestration

1 Introduction and Background

Individual and collaborative learning activities are often combined in educational practice to support social learning experiences [1, 2]. For instance, many widely-used instructional methods (e.g., Think-Pair-Share [3] and Jigsaw [4]) use individual phases to promote productive collaboration. In addition, individual and collaborative modes of learning may have complementary strengths for supporting learning efficiently [5]. For example, collaborative learning offers opportunities for mutual elaboration and co-construction of knowledge, or sense-making; whereas individual learning promotes induction and refinement as learning mechanisms [5].

Given these hypothesized complementary strengths of individual and collaborative learning, it may be fruitful to have students transition *dynamically* between individual and collaborative learning, as the need arises for given students (e.g., when there are diminishing returns in one learning mode at moments where the other might be more effective). Doing so would mean teaming up students in ways that are not fully preplanned, but are instead determined opportunistically based on unfolding learning situations – whether by an instructor or by educational software. Orchestrating this kind of dynamic switching in classrooms has been recognized as a major challenge in teaching practice [6-8]. Prior research has focused on designing tools for supporting teachers in orchestrating either individual (e.g., [9]) or collaborative learning (e.g., [10, 11]) scenarios, or individual and collaborative learning phases on CSCL scripts (e.g., [12]). However, these tools have typically been designed with the assumption that a class of students progresses through instructor- or student-led activities in a pre-planned, relatively synchronized manner [11].

The assumption that transitions are pre-planned breaks down in personalized classrooms such as those where students work with AI-based learning technologies [13]. In practice, teachers dynamically switch between individual and peer tutoring activities. For instance, prior work suggests that during AI-supported class sessions, teachers sometimes orchestrate transitions between individual and collaborative learning on the fly (e.g., by pairing one student to tutor another who may currently be struggling) [14, 15]– although they desire greater support from the AI in doing so [8, 15-17].

In recent years, several projects have begun to explore the design of technologies to support such *human–AI co-orchestration*: the division or sharing of classroom orchestration between different agents in the classroom (e.g. teachers, students, and AI-based systems) [18, 19]. Prior research has explored the design of teacher-centered orchestration tools, which allow teachers to *offload* some decision-making during class, by delegating some student pairing suggestions to the AI system. While offloading was an important goal for teachers, in these studies they also desired a certain amount of *control* over AI systems' decision-making regarding student pairing [8]. That is, they rejected the notion of orchestration systems that operate purely as autonomous agents. Yet, it remains unclear what degree of control would be desirable versus overburdening or distracting to teachers in these contexts [8, 17]. Meanwhile, other work has found that students desire some agency over these decisions as well, and reject the idea of either teachers or AI systems having full control [17]. As this mixed bag of findings indicates, many open questions remain regarding how best to distribute the task of co-orchestrating dynamic transitions between students, teachers and AI systems.

Building upon these prior findings, the current work takes a technology probe approach in live middle school classrooms, to inform the design of co-orchestration support for pairing students and initializing brief, unplanned collaborative interludes during individual work with AI-based tutoring software. Our focus is on understanding how best to distribute control over the dynamic teaming up of students between teacher, students, and the orchestration system. To the best of our knowledge, no prior work has explored how best to support human–AI control over dynamic transitions between individual and collaborative learning in authentic classrooms. Also, this work differs from CSCL literature on dynamic and automated forms of group formation (e.g. [20, 21]) by emphasizing the role of the different actors during the orchestration of dynamic pairing, instead of focusing on the AI technology for pairing up students.

In this paper, we explore design challenges and opportunities for human–AI control over dynamic transitions in the classroom by (1) illustrating student and teacher activity in relation to each pairing policy, (2) exploring students' and teachers' feedback regarding the experience of dynamic switching, and (3) understanding teachers' desires for support and control.

2 Methods

2.1 AI-based Classroom Technologies Used

As platform to explore our vision of dynamically switching students between individual and collaborative learning modes, we used two AI-based tutoring systems, namely, *Lynnette*, which supports individual problem-solving practice, and *APTA 2.0*, which supports mutual peer tutoring, a form of collaborative learning. Both systems support practice in linear equation solving for middle school students. The use of APTA 2.0 helps to address a challenge that has been reported in prior CSCL work [5] and that we have also observed ourselves when teachers spontaneously team up students, namely, that it is difficult for middle-school students (12-14-year-olds) to collaborate effectively. Lynnette provides step-by-step guidance in the form of hints and feedback, as students individually solve equations (e.g. solve for x: $x+3 = 9$). It also keeps track of students' mastery of detailed skill components, as they progress in the problem sets, to support a form of individualized mastery learning. Lynnette is implemented as a rule-based Cognitive Tutor [22] within the CTAT/Tutorshop architecture [23].

To support collaborative learning, we implemented a new version of APTA (Adaptive Peer Tutoring Assistant), developed originally by Walker, Rummel, and Koedinger [24]. This system adaptively coaches one student (the "peer tutor") in tutoring another student (the "tutee") with advice about both tutoring and mathematics. It does so using two rule-based cognitive models, one that captures peer tutoring strategies, one that captures equation-solving knowledge (The latter is shared with Lynnette.) APTA 2.0 supports two different interfaces, one where the tutee (i.e., the student being helped) solves linear equations (Fig. 1 - top), and another through which the student in the peer tutor role monitors the tutee's work and provides guidance (Fig. 1 - bottom). The peer tutor marks their tutee's problem-solving steps as correct or incorrect, accesses hints about equation solving generated by Lynnette, and receives messages from APTA 2.0's coaching model on how to improve tutee's skills and give good advice (e.g., *"Well done! Tutor, do you have a better sense of what your partner is doing?"*, Fig. 1 - bottom). APTA 2.0 connects with Lynnette, which compares the tutee's input to possible correct solutions, so that APTA 2.0 can give the peer tutor feedback on whether their marking of the tutee's steps is correct. APTA 2.0 also presents a chat module where tutees and tutors can communicate during the assignment. For example, tutees can ask for help and tutors can give hints on the current step (Fig. 1, chat component). Chat messages are classified as *help type* (e.g., next-step help, previous-step help, both and not help) [25]. The classification result is then used to feed the coaching model and provide adaptive advice to the tutor. APTA 2.0 assesses students' collaboration skills using its model of collaboration with a variant of Bayesian Knowledge Tracing.

While Lynnette and APTA have been used in prior classroom studies separately and have each separately shown improved learning gains [24, 26, 27], this is the first attempt to combine both AI-systems for dynamically switching between individual and collaborative learning activities.

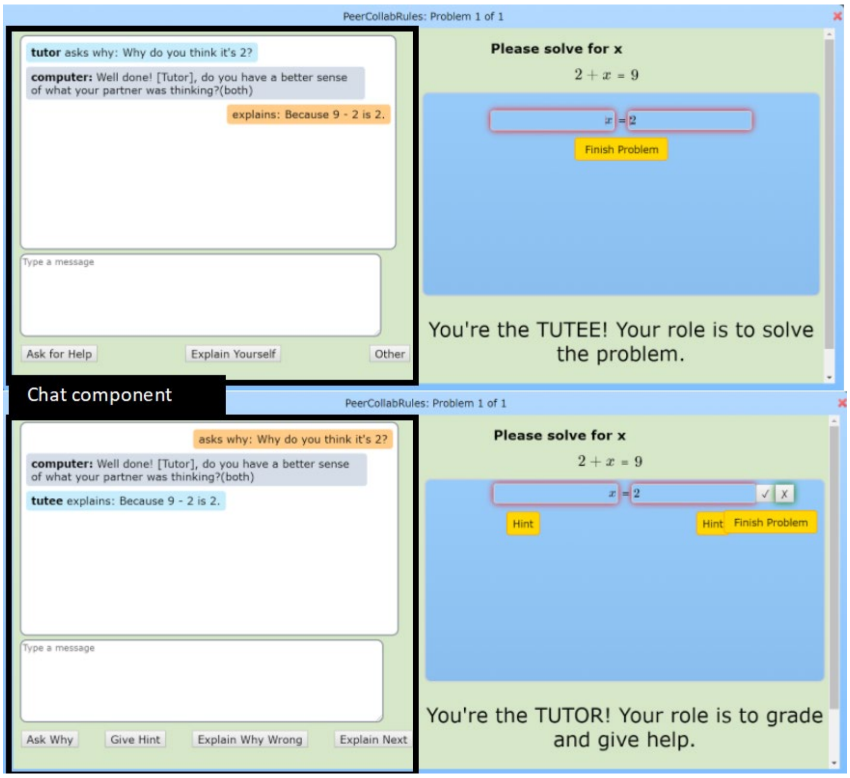

Fig. 1. APTA 2.0 interfaces for students in the tutee (top) and tutor (bottom) role.

2.2 Design, Participants and Procedure

To explore designs of human-AI co-orchestration support for dynamically transitioning between individual and collaborative learning in classrooms, we conducted a technology probe study in middle school classrooms. Our study focused primarily on issues of control (i.e., how can teacher, students, and AI together bring about effective transitions?). Following Hutchinson et al.'s conception of technology probes [28, 29], our goals were to: (1) better understand *how unplanned dynamic pairing plays out* in authentic AI-supported classroom settings, (2) conduct *technical field tests* of an early prototype of a co-orchestration system to support dynamic pairing, and (3) provide teachers and students with the necessary context to provide rich, *experientially-grounded design feedback and ideas* for future human-AI co-orchestration tools.

Three seventh-grade math teachers (1 female, 2 male) from a middle school in a city in the US were recruited. Teachers were asked to use the two AI-based tutoring systems during their normal classroom period. A total of 118 students from six classes (see Table 1), used Lynnette and APTA 2.0 for practicing linear equations. All students had received prior instruction on linear equations before. Thus, the focus was on practice.

Based on prior design explorations with students and teachers [8, 15, 17], we selected and varied the policies for dynamically transitioning between individual and collaborative learning. Each class was randomly assigned to one of three the policies described below. Once the decision to switch given students from one learning mode to another was made, the switch was actuated, in the system, by a (remote) member of the research team. In the future, the steps to do so will be fully automated.

Student pairing policy. Under this policy, *students* were encouraged to request help from a classmate (tutor) if they felt they were stuck on a problem. They could select several peers (based, e.g., on their affinity) by filling in and then submitting a request form. The system (simulated by the remote researcher) initialized a peer tutoring assignment by matching the tutee with their first option listed on the request form. If that option was not available (e.g., because the requested partner was working on another peer tutoring assignment), the system tried to match the tutee with the second option listed, and so on, until fulfilling the help request.

Teacher pairing policy. Under this policy, *teachers* were encouraged to identify a student (tutee) who could potentially benefit from a peer tutoring activity together with a partner (tutor). The teacher could then request that the system (simulated by the remote researcher) pair them up. The teacher could also request to see information about each student's skill mastery in Lynnette, to help with pairing decisions.

AI system pairing policy. An *AI system* (simulated by the remote researcher who constantly monitored students' equation-solving skills as assessed by Lynnette) implemented a policy of teaming up a lower-knowledge student with a higher-knowledge student. The remote observer was instructed to identify students who, for one of the ten skills being monitored (e.g., cancel constant terms), had a probability of knowing – as estimated by Bayesian Knowledge Tracing [30] – below 50% for the tutee and above 75% for the tutor. If the candidate tutor was already paired up with another student, then the observer would choose the next best match, based on any other skill, and ultimately chose, at random, a student who had not been paired up before.

During two regular class periods, each lasting 45 minutes, students performed the following tasks: *First*, they followed a mini-tutorial on how to use Lynnette and APTA 2.0. *Second*, students started to work individually using Lynnette. *Third*, starting after 15 minutes of working individually, students were dynamically teamed up using the pairing policy selected for the given class. Peers (tutee and tutor) were asked to stop their individual work and solve a set of peer tutoring assignments. Following the peer tutoring assignment, which usually involved three problems to solve, students returned to their individual assignments. *Fourth*, students participated in a discussion workshop led by the teacher to discuss their experiences about the pairing policy (e.g., *Did you like to be paired with a peer to solve linear equations? Would you prefer to select your*

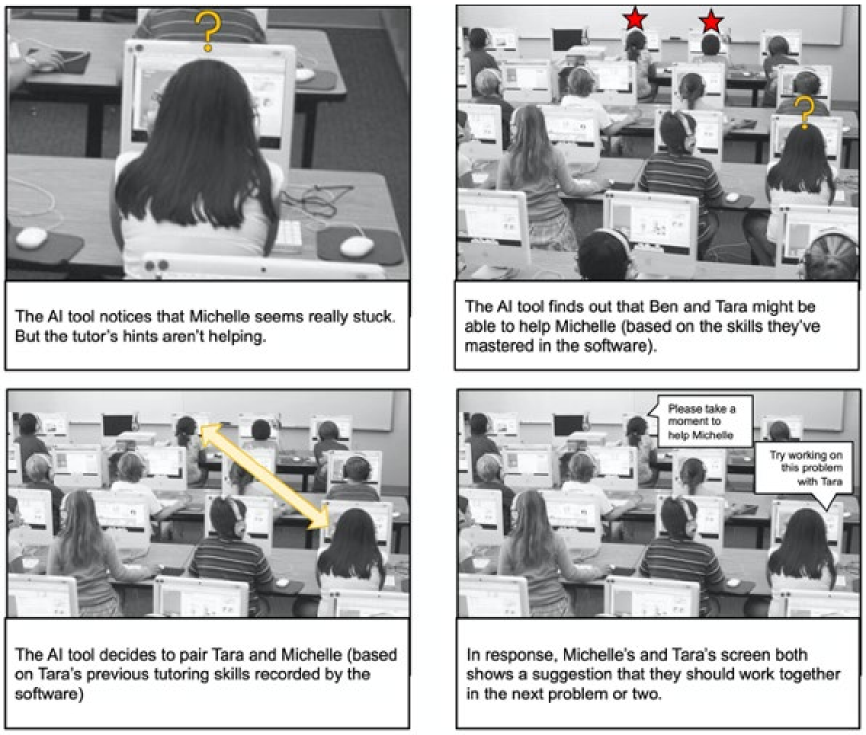

Fig. 2. A storyboard showcasing the pairing policy led by the AI system.

peer? Would you let the teacher, or the system pair you up with someone?) and the overall activity (e.g., *Did you enjoy working with the software?*)

Afterwards, teachers participated in interviews to explore their needs, preferences, and reservations regarding the design of co-orchestration support, building upon their experience during the classroom study. We conducted two semi-structured interviews sessions, each lasting about 30 minutes: one including two teachers (A and B) and the other with one teacher (C). Ideally, all teachers would have experienced all three pairing policies during the classroom part of the study, but that was not feasible for this study. Therefore, to give each teacher an impression of the pairing policy they did not experience, we prepared a set of storyboards representing the three pairing policies (student, teacher, or AI system choice). For example, Fig. 2 depicts a storyboard for the AI system pairing policy. Teachers were asked to review these storyboards, and a researcher led the conversation regarding co-orchestration opportunities (e.g., *Who should have the agency over the pairing policy? Who should accept or reject the initialization of a peer tutoring activity?*) and their preferences.

To preserve students' privacy, we conducted live classroom observations rather than audio/video recording classroom sessions. An observer and a researcher were present during each class period. The observer took observational notes regarding teachers' and students' behaviors using a tool with pre-configured categories (e.g., *teacher explaining* *instructions to the whole classroom*). The researcher took notes during students' workshop sessions and teachers' interviews. All logged data generated by Lynette and APTA were collected in the DataShop repository [31] for further analysis.

3 Analysis, Results and Findings

We aimed to understand design challenges and opportunities for human–AI control over dynamic transitions in the classroom by (1) illustrating student and teacher activity in relation to each pairing policy, (2) exploring students' and teachers' feedback regarding the experience of dynamic switching, and (3) understanding teachers' desires for support and preferences for control.

To address our aims, we analyzed the software log data and classroom observation notes from six classes, along with notes from post-hoc workshop discussions with students and semi-structured interviews with teachers. Qualitative data was analyzed following a content analysis procedure [32]. Quotes of interest were selected by two researchers, and then summarized first, in relation to the specific aims of this study (as presented above) and then, according to each pairing policy. This resulted in a set of insights related to each aim and pairing policy, which are described below.

Table 1. Students distribution for individual and peer tutoring assignments per class and pairing policy.

Pairing policy	Class	Teacher	Individual	Peer tutoring
Student (n=41)	1	A	17	4
	2	B	24	12
Teacher (n=31)	3	B	16	14
	4	B&C	15	12
AI system (n=46)	5	B	26	20
	6	B&C	20	18

3.1 Student and Teacher Activity

Based on the data logs from Lynnette and APTA 2.0 (see Table 1), we found that the *Student pairing policy* yielded fewer peer tutoring assignments (16/41) than the *Teacher pairing policy* (26/31) and the *AI system pairing policy* (38/46).

In the *Student pairing policy*, in which students were encouraged to request to work with a peer when needed, 16 out of 41 students (39.5%) engaged in a peer tutoring assignment. Given the rather large difference in the number of peer tutoring assignments between the two classes who experienced this condition (see Table 1), we analyzed the behavior and characteristics of these two classes from observations and interviews. Only four students from class 1 engaged in peer tutoring assignments possibly because the students in this class were more confident and had more advanced math skills than those in class 2 as mentioned by their teacher (*"class 1 is a high-achieving classroom",* Teacher A). As for students from class 2, half of students worked on a peer tutoring assignment, a low number compared to the other two conditions (e.g., 14 out of 16 students from class 3 worked on a peer tutoring activity). Evidently, not all students are equally motivated to work on collaborative learning activities. For instance, the majority of students from class 2 stated that *"they would prefer working alone."*

In the *Teacher pairing policy*, the teacher was the instigator of the dynamic pairing. Table 1 shows that 26 out of 31 students (83.9%) of students in this condition engaged in a peer tutoring activity, meaning that the teacher was able to make decisions about how to pair up students by herself (14 out of 16 students), and by requesting information from students' skills, which were retrieved from Lynnette (12 out of 15 students). While a deep data analysis would be needed to understand if this pairing strategy led to more effective peer tutoring activities than other pairing strategies, these results suggest that retrieving assessments of students' mastery of math skills from the AI system could potentially help teachers to make informed decisions for pairing up students.

In the *AI system pairing policy*, 38 out of 41 (92.7%) students did peer tutoring assignments. Of these peer tutoring partnerships, 78.9% (30 of 38) were chosen based on the "ideal" matching criterion, namely, that for at least one of the equation-solving skills targeted in the instruction, the tutee's mastery level was below 50% and the tutor's mastery level was above 75%. In addition, 15.8% (6 of 38) of peer tutoring partnerships were chosen based on the next best match, which required at least one that the peer tutor had higher mastery of at least one skill. Finally, 5.23% (2 of 38) were initialized by randomly selecting a peer tutor due to a lack of good candidates. These results suggest that the *AI system pairing policy* seems to be feasible for teaming up students, or for suggesting students to be teamed up when teachers want control.

3.2 Student and Teacher Feedback

Student pairing policy. All students from class 1 (high-achieving classroom) and some students in class 2 (large classroom) liked being able to choose a classmate and not being paired by the teacher. A student from class 1 stated that he would prefer *"to choose someone I can work with better"* based on their affinity. However, another student argued that he would prefer to *"work with someone with a higher skill level"*. The majority of students in class 2 stated that they would prefer to work alone, and some students thought it would be better to ask for help from the teacher instead of a classmate.

Teachers' views did not align with students' preferences. Teachers commented that they would prefer to have more control over the dynamic transitions. For example, they stated that they would like to *"have some control over the pairing"* and *"override students' pairing,"* arguing that *"some kids don't work well might choose those who'll just give answers or chat about something else."* Thus, one reason teachers may prefer to have the final say over the pairing is a concern that some students may not collaborate effectively, and that students' collaboration skill should be a factor in teachers' preferences for teaming up students. Estimates of how well students mastered the collaboration skills could be retrieved from APTA's coaching model [24]. It is clear as well that teachers prefer to limit non-math related chat interactions on the part of students. This concern appears well-founded. Findings from prior work [24] and chat interactions from the current study indicate that it is very common for students to exchange off-topic messages through chat. For instance, for all three pairing policies we found a high percentage (84.9%) of tutors' messages related to *off-topic* chat entries.

Teacher pairing policy. The majority of students from class 3 and roughly half of the students from class 4 (small classroom) approved of letting the teacher make pairing decisions. One student stated*: "she [the teacher] knows who is good and who is bad,"* noting that the teacher could use her prior knowledge of the students' skills to get a productive peer tutoring activity. However, some students from class 3 and class 4 stated that they would like to choose a classmate to become a tutor *"depending on the problem [they] are working on,"* meaning that they would expect to be helped by a friend or someone in the class only if the problem is not too difficult, and otherwise would prefer to receive help from the teacher rather than a peer. Although most students agreed with the pairing decisions made by their teacher, they also recommended other pairing policies. For example, one student mentioned that she would prefer *"to get a randomized partner because he can get someone new every time."* However, another student raised the concern that, with randomized partners, *"it could be possible to get someone who cannot help you with the problem."* Following up on this idea, other students suggested that another pairing policy to match tutees and tutors would be *"based on a [tutor] skill"* or *"a qualification of becoming a tutor,"* and only students who get this skill should be recommended for tutoring other students. These comments from students are in line with the implications of some of the teachers' comments discussed above. They raise the interesting idea of a pairing policy that would take into account APTA's assessment of a student's peer tutoring skills (as mentioned, APTA 2.0 uses its model of tutoring skill to assess individual students' skill in this area).

Teachers agreed that the teacher pairing policy might be more beneficial for students, as the teacher would choose someone who *"they can focus better."* However, teachers raised some concerns about the orchestration load from teachers' side, as one teacher expressed that *"at some point, matching and monitoring individual and peer tutoring activities would be bothersome."*

AI system pairing policy. The majority of students from class 5 (large classroom) agreed to be matched by the system, while most from class 6 (low-achieving classroom) had a contrasting view, expressing that they would prefer to be able to choose a classmate to become their tutor. Comments from students also indicated that they felt surprised to learn who their peer was. One student said that *"his peer was a classmate who does not talk to him often."* Similar to previous comments about the other pairing policies, students indicated that the matching of tutees and tutors should be based on skill levels. As one student suggested: *"if the skill [represented as a horizontal bar in the interface] is long, that person should be a tutor."*

Teachers liked the idea of having an AI system pair up students. However, they pointed out they should have some control over the AI system's decisions. One teacher mentioned that *"he must be able to override the system's matching decisions."* Teachers also suggested that the system should be able to have some constraints for matching certain students, based, presumably, on their knowledge about students' characteristics and knowledge. For example, a teacher stated that *"he would trust the system",* as long as it has some constraints such as *"never putting these two kids (e.g., Sally and Molly) together because I know they don't work together well."* Furthermore, teachers indicated that the AI system could potentially suggest best matches to teachers, based on students' skills, along with the teacher having the ability to accept or reject a matching suggestion.

3.3 Further Understanding Teachers' Desires for Support and Control

When teachers were asked about their desires for support and control of the dynamic transitions, they all envisioned *hybrid* forms of control shared between students, teachers, and AI systems. Additionally, teachers indicated that the control of the dynamic transitions should be tailored to class and individual student characteristics, noting that the orchestration tool should *"preserve flexibility,"* because *"different classes have different dynamics and skills."* This view is supported through the results presented in Section 3.1 and Section 3.2. We explain this view according to three differentiated classroom characteristics that emerged from teachers' interviews.

High-achieving classrooms. There was a high-achieving classroom in the study (Class 1): as mentioned, only 4 of 17 students from this class worked on peer tutoring activities (see Table 1). Teacher A stated that most of the students from his class were taking advanced math classes. Teachers suggested that for this particular class, the pairing policy could be co-orchestrated by *students* and *teachers*, such that students would have some agency with respect to the pairing policy, and the teacher would have the option to accept or reject them in the orchestration tool.

Low-achieving classrooms. In the low-achieving classroom (Class 6), 18 out of 20 students worked in peer tutoring activities (see Table 1). Their teacher (Teacher B) stated that there were several low-achieving students and that usually, the classroom dynamics for this class are different from others (e.g., large classes like class 2). Teachers indicated that for this class, orchestration could be divided between the *AI system* and *teachers*. For example, teacher C indicated that she would let the AI system match students according to students' skills, as long as the system is *"able to restrict some matchings"* depending on students' characteristics (e.g., affinity). All teachers also suggested that the teacher should be able to accept and reject these pairings. However, Teachers B and C opined that, at the same time, *the teacher should not become a bottleneck*: if the teacher is busy helping another student or doing any other classroom duty and does not have the time to accept or reject the pairing, the system should be able to proceed and initialize the peer tutoring activity.

Classroom sizes. During interviews, teachers also noted that their ability to share control over dynamic pairings with the AI system is constrained by the size of the class – a factor that has rarely been explored in prior work on co-orchestration support. For instance, one teacher stated that complete teacher control over the dynamic pairing would be feasible in a *small class*, if *"most students could work in individual assignments,"* and only a small number of students could benefit from a peer tutoring activity. By contrast, in cases where teachers need to orchestrate larger classes, they would prefer to maximize the support from *AI systems* to offload some of the orchestration tasks, as *"it would take much time to control and monitor too many students at once."* Thus, in larger classes teachers were open to exploring the option of giving most of the agency to the AI system to monitor and suggest pairing opportunities. Teachers clearly saw value in *sharing* control over these dynamic transitions with an *AI system*.

4 Discussion and Future Directions

Motivated by prior findings that suggest that students benefit from alternating *dynamically* between individual and collaborative activities, where such alternations are not pre-planned, but are instigated on the fly, as the need arises, we aimed to understand how to divide or share control between teachers, students, and AI systems over students' transition between the two types of learning activities.

In line with prior research on co-orchestration, findings from the current study suggest a need for a form of *hybrid control* shared among students, teachers, and AI systems [7, 18, 33]. As in prior investigations, students and teachers expressed differing preferences regarding how control over peer tutor selection should be distributed among teachers, students, and the AI [17]. Students' feedback regarding pairing policies indicated that both the *Student* and *Teacher pairing policies* were well received. However, students were hesitant regarding the *AI system policy*. In contrast, teachers showed a greater affinity towards the *Teacher* and *AI system pairing policies*. In addition, the current work suggests a need for the distribution of control to be sensitive to classroom characteristics (e.g., class size and distribution of ability among students) and dynamics (e.g., potential for teacher overload). For instance, our findings suggest that for high achieving classrooms, in line with prior findings, *students* and the *teacher* could share control over the pairing by allowing students to select their partner while enabling the teacher to have oversight over these selections [17, 33]. As for low achieving classrooms, our findings point to sharing of control mostly between the *teacher* and the *AI system*, with the AI system matching students according to their skill mastery or peer tutoring (coaching) ability, while also taking teacher-specified constraints into account (e.g., preventing certain pairs of students from working together, or preventing any given student from serving as a peer tutor too often).

Moving beyond prior work on the design of co-orchestration support, the current work points to needs for adaptivity and adaptability for different classroom contexts, teacher preferences, and students' prior knowledge [7]. One example of *adaptability* in a co-orchestration system could be enabling *teachers* to select the best pairing policy based on their particular goals, needs, and classroom dynamics. In a small class, or one in which few students are struggling or in need of additional help, teachers may allow students, at some point during class, to choose which peers they would like to work with – perhaps supported by the AI system. By contrast, in a large class, or one where many students go through a rough stretch, teachers might choose to have the AI system take more control over the pairing decisions, within constraints preconfigured by them [cf. 8], and with the option of approving, adjusting, or vetoing the system suggestion. In addition to adaptability, our findings also suggest promise for co-orchestration systems that are *adaptive* to particular teachers' and classrooms' needs. For example, a co-orchestration system might detect the class size or the teacher's current workload and in turn adjust how it balances control across teachers, students, and the AI. When the teacher has a high workload, the system could intervene by automatically assuming greater control over orchestration to support more fluid transitions (e.g., by ensuring the teacher is not a bottleneck for pairing decisions).

While the above examples illustrate how co-orchestration systems can be designed to respond to diverse classroom situations, several open research and design questions remain. Further research is needed to understand the *right balance* of teacher, student, and AI control over pairing decisions, which our findings suggest may need to be understood as *adaptive* to different classroom contexts. In particular, our findings point to a need to explore the design space of context-adaptive pairing policies. Future studies, conducted in a larger number of classrooms, may shed light on how a broader range of classroom characteristics and student's factors could affect dynamic transitions.

5 Conclusions

This exploratory study aimed to address the design of human-AI co-orchestration systems that meet the complexity of authentic classrooms. To the best of our knowledge, this is the first classroom field study to explore human–AI control over dynamic transitions between individual and collaborative learning. This study yields experientially-grounded feedback from teachers and students, to inform the design of co-orchestration support for dynamic transitions. Moving beyond prior work in this area, which has offered *general* design recommendations for "average" classroom contexts, this study surfaced *context-dependent* needs for the design of human-AI co-orchestration support. General design guidelines for orchestration technologies have emphasized the need to carefully consider classroom context and students' characteristics. Yet little research has explored how this goal might be achieved in contexts where orchestration is distributed among humans and AI systems.

In sum, this work contributes to the emerging literature on human–AI co-orchestration, pointing to needs for further research on how particular orchestration tasks can best be balanced between teachers, students, and AI systems, and on how the ideal balance may depend on classroom contextual factors. In turn, the design of new co-orchestration supports may facilitate complex yet powerful classroom scenarios, which would otherwise be difficult or impractical to implement.

Acknowledgements This work was supported by grants IES R305A180301 and NSF (IIS) 1822861. The opinions expressed are those of the authors and do not represent the views of the funding agencies. Special thanks to all participating teachers and students.

References

1. Dillenbourg, P.: Over-scripting CSCL: The risks of blending collaborative learning with instructional design. (2002)
2. Dillenbourg, P.: Framework for integrated learning. (2004)
3. Kothiyal, A., Murthy, S., Iyer, S.: Think-pair-share in a Large CS1 Class: Does Learning Really Happen? Conference on Innovation & Technology in Computer science Education, pp. 51-56. Association for Computing Machinery, Uppsala, Sweden (2014)
4. Aronson, E.: The jigsaw classroom. Sage (1978)
5. Mullins, D., Rummel, N., Spada, H.: Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. Intl. Journal of Computer-Supported Collaborative Learning 6, 421-443 (2011)
6. Kaendler, C., Wiedmann, M., Rummel, N., Spada, H.: Teacher competencies for the implementation of collaborative learning in the classroom: A framework and research review. Educational Psychology Review 27, 505-536 (2015)
7. Dillenbourg, P., Nussbaum, M., Dimitriadis, Y., Roschelle, J.: Design for classroom orchestration. Computers & Education 69, 485-492 (2013)
8. Olsen, J.: Orchestrating Combined Collaborative and Individual Learning in the Classroom. Carnegie Mellon University (2017)
9. Alavi, H.S., Dillenbourg, P., Kaplan, F.: Distributed awareness for class orchestration. European Conference on Technology Enhanced Learning, pp. 211-225. Springer, Nice, France (2009)
10. Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J.: MTFeedback: Providing Notifications to Enhance Teacher Awareness of Small Group Work in the Classroom. IEEE Transactions on Learning Technologies 8, 187-200 (2015)
11. van Leeuwen, A., Rummel, N.: Orchestration tools to support the teacher during student collaboration: a review. Unterrichtswissenschaft 47, 143-158 (2019)
12. Tissenbaum, M., Slotta, J.: Supporting classroom orchestration with real-time feedback: A role for teacher dashboards and real-time agents. Intl. Journal of Computer-Supported Collaborative Learning 14, 325-351 (2019)
13. Ritter, S., Yudelson, M., Fancsali, S.E., Berman, S.R.: How Mastery Learning Works at Scale. Third Conference on Learning @ Scale, pp. 71-79. Association for Computing Machinery}, Edinburgh, Scotland, UK (2016)
14. Schofield, J.W., Eurich-Fulcer, R., Britt, C.L.: Teachers, computer tutors, and teaching: The artificially intelligent tutor as an agent for classroom change. American Educational Research Journal 31, 579-607 (1994)
15. Holstein, K., McLaren, B.M., Aleven, V.: Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. Journal of Learning Analytics 6, 27-52 (2019)
16. Holstein, K., McLaren, B.M., Aleven, V.: Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. Intl. Learning Analytics & Knowledge Conference. ACM, Vancouver, British Columbia, Canada (2017)
17. Holstein, K., McLaren, B.M., Aleven, V.: Designing for complementarity: Teacher and student needs for orchestration support in ai-enhanced classrooms. Intl. Conference on Artificial Intelligence in Education, pp. 157-171. Springer, Chicago, IL, USA (2019)

18. Prieto, L.P., Dlab, M.H., Gutiérrez, I., Abdulwahed, M., Balid, W.: Orchestrating technology enhanced learning: a literature review and a conceptual framework. Intl. Journal of Technology Enhanced Learning 3, 583 (2011)
19. van Leeuwen, A., Rummel, N., Van Gog, T.: What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? Intl. Journal of Computer-Supported Collaborative Learning 14, 261-289 (2019)
20. Amarasinghe, I., Hernández-Leo, D., Jonsson, A.: Intelligent group formation in computer supported collaborative learning scripts. 17th Intl. Conference on Advanced Learning Technologies (ICALT), pp. 201-203. IEEE (2017)
21. Srba, I., Bielikova, M.: Dynamic group formation as an approach to collaborative learning support. IEEE Transactions on Learning Technologies 8, 173-186 (2014)
22. Long, Y., Aleven, V.: Mastery-oriented shared student/system control over problem selection in a linear equation tutor. Intl. Conference on Intelligent Tutoring Systems, vol. 9684, pp. 90-100. Springer, Zagreb, Croatia (2016)
23. Aleven, V., McLaren, B.M., Sewall, J., Van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., Koedinger, K.R.: Example-tracing tutors: Intelligent tutor development for non-programmers. Intl. Journal of Artificial Intelligence in Education 26, 224-269 (2016)
24. Walker, E., Rummel, N., Koedinger, K.R.: Adaptive intelligent support to improve peer tutoring in algebra. Intl. Journal of Artificial Intelligence in Education 24, 33-61 (2014)
25. Walker, E., Walker, S., Rummel, N., Koedinger, K.R.: Using problem-solving context to assess help quality in computer-mediated peer tutoring. Intl. Conference on Intelligent Tutoring Systems, pp. 145-155. Springer (2010)
26. Long, Y., Aleven, V.: Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. User Modeling and User-Adapted Interaction 27, 55-88 (2017)
27. Waalkens, M., Aleven, V., Taatgen, N.: Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. Computers & Education 60, 159-171 (2013)
28. Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B.B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H.: Technology Probes: Inspiring Design for and with Families. SIGCHI Conference on Human Factors in Computing Systems, pp. 17-24. ACM, Ft. Lauderdale, Florida, USA (2003)
29. Quintana, R., Quintana, C., Madeira, C., Slotta, J.D.: Keeping Watch: Exploring Wearable Technology Designs for K-12 Teachers. CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2272-2278. Association for Computing Machinery, San Jose, California, USA (2016)
30. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction 4, 253-278 (1994)
31. Koedinger, K.R., Baker, R.S., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.: A data repository for the EDM community: The PSLC DataShop. Handbook of educational data mining 43, 43-56 (2010)
32. Hsieh, H.-F., Shannon, S.E.: Three approaches to qualitative content analysis. Qualitative health research 15, 1277-1288 (2005)
33. Sharples, M.: Shared orchestration within and beyond the classroom. Computers & Education 69, 504-506 (2013)