Surveying Teachers' Preferences and Boundaries Regarding Human-AI Control in Dynamic Pairing of Students for Collaborative Learning

Kexin Bella Yang[1](✉), LuEttaMae Lawrence[1], Vanessa Echeverria[2], Boyuan Guo[1], Nikol Rummel[3], and Vincent Aleven[1]

[1] Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, USA
{kexiny,llawrenc,boyuang,va0e}@andrew.cmu.edu
[2] Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Ecuador
vanechev@espol.edu.ec
[3] Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany
nikol.rummel@rub.de

Abstract. Orchestration tools may support K-12 teachers in facilitating student learning, especially when designed to address classroom stakeholders' needs. Our previous work revealed a need for human-AI shared control when dynamically pairing students for collaborative learning in the classroom, but offered limited guidance on the role each agent should take. In this study, we designed storyboards for scenarios where teachers, students and AI *co-orchestrate dynamic pairing* when using AI-based adaptive math software for individual and collaborative learning. We surveyed 54 math teachers on their co-orchestration preferences. We found that teachers would like to share control with the AI to lessen their orchestration load. As well, they would like to have the AI propose student pairs with explanations, and identify risky proposed pairings. However, teachers are hesitant to let the AI auto-pair students even if they are busy, and are less inclined to let AI override teacher-proposed pairing. Our study contributes to teachers' needs, preference, and boundaries for how they want to share the task and control of student pairing with the AI and students, and design implications in human-AI co-orchestration tools.

Keywords: Classroom · Human-AI collaboration · CSCL · HCI · Design orchestration tools

1 Introduction

While teachers need to facilitate students' collaborative activities and monitor their progress, these activities can be cognitively demanding for teachers [1–4]. Orchestration

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-86436-1_20) contains supplementary material, which is available to authorized users.

T. De Laet et al. (Eds.): EC-TEL 2021, LNCS 12884, pp. 260–274, 2021.
https://doi.org/10.1007/978-3-030-86436-1_20

broadly refers to the planning and real-time management of learning activities in the classroom [5]. Various orchestration tools and prototypes have been designed to help lessen teachers' load of class management or allow them to focus on teaching and helping students [3, 4, 6–8]. Recently, much work focuses on *"co-orchestration,"* referring to technology that is explicitly designed so that the orchestration responsibilities are shared across multiple agents (e.g., teachers, students, and AI systems) [4, 5, 7].

Much research has focused on designing tools to support teachers orchestrating either individual (e.g. [9]) or collaborative learning scenarios [2, 10]. These tools have typically been designed with the assumption that students progress through activities in a pre-planned, relatively synchronized manner. By contrast, little work has focused on supporting *fluid social transitions* (with recent exceptions [3, 6, 11]). *Fluid social transitions,* as defined by Olsen et al. [3], refer to transitions between classroom activities that "occur asynchronously between students - not all at the same time for everyone in the class." Different from the static, planned transitions, fluid social transitions may be needed in technology-enhanced classrooms, to support students to flexibly move between activities at a pace that suits their specific circumstances and knowledge level. It is a hypothesis that technologies that support fluid transitions can better support students' learning [3]. To be able to test this hypothesis, the high-level goal for the current study is to design a co-orchestration tool to help teachers easily and effectively manage fluid social transitions in class. Specifically, we focused on dynamic pairing, which means teaming up students opportunistically based on unfolding learning situations [3, 11].

To design effective tools that can help K-12 teachers co-orchestrate their students' learning activities in classrooms, it is critical that the tools "support the needs and respect the boundaries of both teachers and students" [12]. Researchers have begun exploring questions such as: How should orchestration responsibilities be divided among different classroom stakeholders? Who should be accountable for particular instructional decisions? How should such hybrid control adapt to context and learning scenarios [5, 11, 12]? Our previous work reveals that classroom stakeholders have nuanced preferences regarding the co-orchestration of classroom activities. From design research with K-12 teachers, researchers found "a delicate balance between automation and respecting teachers' autonomy", and that "over-automation risks threatening teachers' authority in class and flexibility to set their own goals, yet under-automation may burden teachers with tasks they'd rather not perform" [12]. Similarly, Olsen et al. found from design research with primary school teachers that instructors prefer to "maintain an elevated position above AI systems and need to have some degree of accountability and control" [3]. Echeverria et al. found from a Wizard-of-Oz technology probe in K-12 classrooms a need for hybrid control between students, teachers, and AI systems over dynamic transitions from individual to collaborative learning. They also found a need for such hybrid control to be adaptable to classroom contexts, such as class size and students' prior knowledge.

This study investigates the following questions: *In co-orchestrating dynamic student pairing, how do teachers want to share control with AI systems and students, regarding proposing, evaluating, and deciding pairings?* (RQ1) *What criteria do teachers prefer when dynamically pairing students?* (RQ2) In the current study, we designed storyboards of possible scenarios of human-AI co-orchestration of dynamic pairing. We surveyed

54 math teachers' preferences on them, in the context of using adaptive AI-tutoring software for individual and collaborative K-12 math learning.

2 Methods

2.1 Study Design

The overall goal of the research project, that the current study is part of, is to design a co-orchestration tool that helps teachers manage fluid transitions back-and-forth between individual and collaborative learning in the classroom. This study concerns finding teachers' needs and preferences on one key aspect in managing fluid transitions: dynamically teaming up students (i.e., dynamic pairing). Dynamic pairing may not happen at the same time for every student, and managing the process in real-time may be overwhelming for teachers.

Many design-based research activities, such as user interviews, can require significant time input from researchers, often resulting in a relatively smaller participant pool. Surveys, while being easily scalable, are often hard to adequately convey the nuanced designs to the participants. To concisely communicate potential designs to a larger user population, we incorporated storyboards into a survey. We created storyboards that illustrate different co-orchestration tool features in the context of different dynamic scenarios. We surveyed teachers' opinions on these scenarios, to make sure the tool we design can be aligned with teachers' needs and preferences. These scenarios take context in two intelligent tutoring systems (ITSs) that support learning of middle school math algebra, specifically, equation solving. The individual ITS, Lynnette, has been proven to improve students' equation solving skills in several classroom studies (e.g., [13]). The collaborative ITS, APTA, extends Lynnette's functionality to support reciprocal peer tutoring.

Using storyboards is a standard method in human-centered design. Storyboards show how users interact with different versions of a proposed system in specific contexts [14]. According to Davidoff et al., in the need validation process, storyboards may "help designers prioritize user needs, more clearly map spaces for innovation, and use that focus to narrow the design space for potential applications" [14].

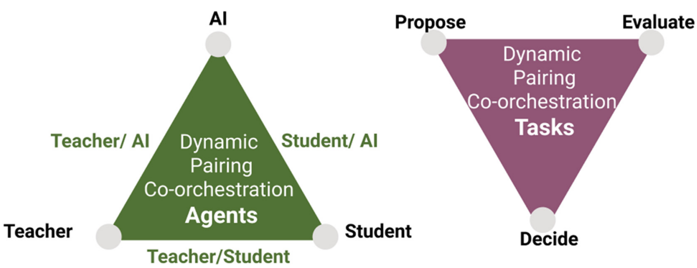

Fig. 1. Agents (left) and Tasks (right) in Human-AI Orchestration of Dynamic Student Pairing

Co-orchestration Scenarios and Survey Design. To inform the design space, we reviewed literature at the intersection of orchestration tools, dynamic pairing, fluid social transitions, and computer-supported collaborative learning (CSCL) [3, 11, 15–17]. As shown in Fig. 1, for co-orchestrating dynamic pairing, there are three main *agents* and three main *tasks*. The three main agents are teacher, AI, and student (Fig. 1, left), and the three main tasks are how the student pairings are proposed, evaluated, and decided (Fig. 1, right). The design space for dynamic pairing co-orchestration is about which *agent(s)* have control over each *task*. Two authors collaboratively brainstormed 22 design concepts for potential co-orchestration scenarios. We used our past research on teachers' needs in pairing co-orchestration [3, 11] to prioritize scenarios that teachers may find more useful. We also included scenarios that push social boundaries or maybe controversial among teachers, as it may help to uncover where these boundaries lie [14]. We went through four rounds of clustering and refining the 22 design concepts, and finalized five co-orchestration scenarios, described in Table 1 (first column).

Based on these five scenarios, we designed a survey that had three sections: 1) teachers' demographics and teaching experience, 2) five co-orchestration scenarios and related questions, and 3) teachers' general preferences on pairing co-orchestration. The five scenarios were presented in a random sequence for every participant, to reduce bias and carry-over effects. Each scenario has 1) a three-panel storyboard with a short title with a simple visual and description (e.g., Fig. 2), 2) a seven-point Likert scale question asking how likely it is that the teacher would use the technology shown in the storyboard in their classroom, 3) a follow up open-ended question asking why or why not, 4) an open-ended question asking what improvements or changes teachers would want to make to the scenario, 5) 1–2 focused seven-point Likert question specific to the main design element in the scenario. In addition, some scenarios had 6) 1–2 multiple choice probing questions. The complete survey questions can be found in supplementary materials.

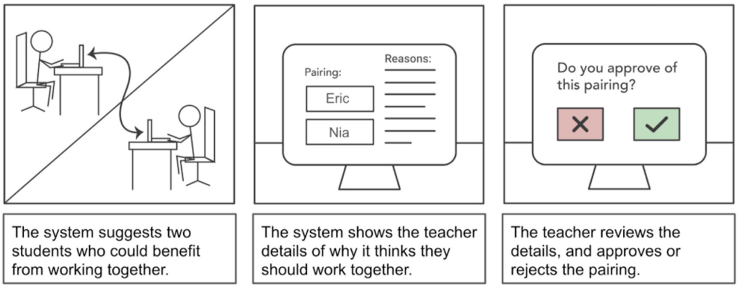

Fig. 2. Storyboard Example (Co-orchestration Scenario 3)

Procedure. The survey was hosted online using Qualtrics and available for seven days. Teachers were introduced to the context and asked to consent to the research. Before large-scale deployment, researchers conducted 8 pilot sessions including 2 think-aloud studies, with math teachers and researchers with 5+ years experience in classroom orchestration. Based on results from the pilot, researchers iteratively refined the survey and storyboards to ensure clarity.

Participants. 54 math teachers (36 females, 18 males) were recruited from math teacher groups on social media. They were asked to complete the 20 to 30-min survey and paid $10 USD Amazon Gift Card. Most teachers (75%) taught in middle school and had taught math for 4 or more years (88%). Participants were mostly white (80%) and mainly taught in the United States of America (96%). At the time, 52% were teaching fully online or remote settings, 41% were in a hybrid mode of online and in-person teaching, and only 2 teachers (4%) were teaching fully in-person. Most teachers (85%) used collaborative learning activities in their classrooms for 50% of the time or more. Other demographics can be found in the supplementary materials.

Table 1. Co-orchestration Scenario Descriptions and Teachers' Stated Likeliness of Use

Scenario titles and descriptions	Co-orchestration roles	Stated likeliness of use	
		M (SD)[a]	Median
S1. Teacher Informs Automated Pairing: Teacher entered information for the AI to use when making student pairings	*Proposes*: Teacher and System *Evaluate*: N/A *Decides*: AI	4.96 (1.58)	5
S2. AI Auto-Pairs When Teachers are Busy: AI paired students up on its own because the teacher was busy and in "Do not disturb" mode	*Proposes*: AI *Evaluate*: N/A *Decides*: AI	4.70 (1.59)	5
S3. AI Explains Pairing Suggestions: AI paired students up, explained to the teacher why it paired them up and asked the teacher to approve or reject the pairing	*Proposes*: AI *Evaluate*: Teacher *Decides*: Teacher	5.89 (1.13)	6
S4. AI Identifies Risky Pairs: Students proposed a partner, the AI reviewed the pairing and notified teachers of potentially risky pairings for teachers to approve or reject	*Proposes*: Student *Evaluate*: AI+ Teacher *Decides*: Teacher	5.33 (1.41)	6
S5. AI Reviews and Decides: Teachers paired students up, the AI evaluated the pairings, and changed risky ones without notifying teachers	*Proposes*: Teacher *Evaluate*: AI alone *Decides*: AI	4.00 (1.76)	4

[a]Likert Scale Labels: 1 - Definitely No, 2 - Very Unlikely, 3 - Probably not, 4 - Neither likely nor unlikely, 5 - Probably yes, 6 - Very likely, 7 - Definitely yes

2.2 Data Analysis Approach

To ensure response quality, we manually reviewed survey responses, and filtered out duplicate or invalid ones (e.g., coming from the same IP address).

For the *quantitative analysis,* we computed the mean and SD for each seven-point Likert scale questions (e.g., how likely teachers thought it was that they would use the technology depicted in the storyboard), and analyzed teachers' responses to each of the multiple-choice probing questions (e.g., what student information the system should consider in dynamic pairing). For the *qualitative analysis,* we analyzed a total of 540 open-ended teacher responses to the two open-ended questions: 1) why they would or would not use the technology in their classroom, and 2) what they wish to see changed or improved. We conducted iterative affinity diagramming [18], where two members of the research team grouped and regrouped individual pieces of data (i.e., raw teachers' response to the two questions above) to find common themes representing teachers' opinions. These teachers' comments were iteratively clustered into 117 first level themes and 63 second-level themes.

To synthesize findings, we laid out all the analysis results *within each scenario*, which allowed us to see, from quantitative data, how teachers' *preferences* on design elements in the co-orchestration tool were distributed, and *why they felt this way* from qualitative data. We then organized the quantitative results and qualitative themes *across different scenarios* according to research questions. Based on this organization, we then formulated insights regarding teachers' co-orchestration preferences, attending both quantitative distribution and qualitative sentiments.

3 Results

In Sect. 3.1, we report quantitative and qualitative results, organized by *scenario*, regarding teachers' needs and preferences for co-orchestration of dynamic pairing. In Sect. 3.2, we report teachers' opinions on hybrid control in proposing and deciding who to pair. The supplementary materials contain complete survey questions and detailed statistics for all Likert scale questions responses.

3.1 Quantitative and Qualitative Results for Each Scenario

Scenario 1 Results. In the first scenario (S1), teachers can enter information about each individual student, which the AI uses to make student pairings later that week. S1 was third-highest regarding teachers' stated likeliness of use ($M = 4.96$, $SD = 1.58$). The survey then asked whether teachers wanted to *spend the time to enter student information* that the AI could use to pair up students. In answer to this question, 37% of teachers chose "very likely" or "definitely yes" in the seven-point Likert scale ($M = 4.65$, $SD = 1.85$). In response to the question asking whether different students should be paired using the *same or different criteria*, the majority of teachers (70%) wanted *different* (personalized) pairing criteria for each student, 26% wanted to have the *same* pairing criteria for all students, and 4% chose "other". When asked what *factors* the system should consider when pairing students, and given a list from which they could select multiple factors, almost all teachers thought the system should consider *students' level of knowledge mastery* (98%). Other factors included students' error rate (76%), students' pairing history (76%), students' personality (66%), number of collaborative sessions students already did (64%), students' friendship/ relationship (52%) and gender (34%).

Qualitative analysis from scenario 1 showed that teachers generally like that the technology could pair students in a data-driven way and that it may increase their work efficiency and reduce orchestration load ($N = 7$)[1]. They also liked the fact that such a system may allow them more work flexibility because they could front-load the preparation work for students pairing prior to class sessions ($N = 6$). However, some teachers were hesitant to use such technology in their classroom, mainly because they were concerned that manually entering students' information would be time-consuming ($N = 12$). Such concerns were amplified for teachers with larger classes: *"I would not spend even 5 min entering a bunch of data about individual students to be used in pairing later. I have 130 students - I can't imagine how time-consuming that would be"* (T3). Some teachers mentioned privacy concerns ($N = 2$), e.g., "*I am also concerned about the type of data the system will ask for in regards to students. Personality? Friendships? I do not think this is appropriate information to enter into an educational software system.*" (T3) Accordingly, teachers wished for a more efficient, less burdensome way to achieve the same effect as manually entering the information ($N = 7$). For example, some teachers prefer "*the system to collect the data rather than enter it myself.*" (T7) or "*students taking a self-assessment*" (T19).

Scenario 2 Results. In the second scenario (S2), the AI system auto-pairs students when the teacher turns on a "Do not disturb" mode in the orchestration tool, to signal they are busy helping a student. In other words, in this scenario, the AI sometimes has full autonomy over pairing decisions. S2 has the *second-lowest* stated likeliness of use among teachers ($M = 4.7$, $SD = 1.59$). There was substantial variability in teachers' answers to the question of whether teachers think the system should pair students without asking teachers' approval when they are busy. Only 37% of teachers chose "agree" or "strongly agree" ($M = 4.35$, $SD = 1.78$). The survey also asked *when* teachers preferred the AI to pair students up without asking for their approval. Most (57%) teachers chose *when helping other students*, 22% chose when they are *off work*, and 9 teachers (17%) chose *never*.

From the qualitative analysis, we found teachers generally think that auto-pairing by the co-orchestration tool when the teacher is busy could help *reduce interference and distraction* in teachers' work (e.g., *"It can effectively reduce the disturbing information for teachers"* [T46]). They also opined that it could improve their work efficiency ($N = 19$). They also liked the fact that such a system may help students to get help sooner, (e.g., *"I like the fact it can partner students without teacher action"* [T41]), engage better in class, and learn collaboratively ($N = 8$). However, many teachers had reservations about a possible auto-pair feature ($N = 32$). Specifically, some teachers did not want the system to have such a high degree of control over the pairing process and wished to be able to review the pairing, and change or override the system's pairing decisions when needed ($N = 26$). As one teacher described "*If a computer can pair students at random, then what is the point of a teacher being present? This seems to undermine the authority of a teacher in the classroom"* (T54). Some other teachers ($N = 7$) thought this feature "*seems needlessly complicated*" and preferred that the system *"just give[s]*

[1] N refers to the number of teachers' comments.

a notification to the side that's unobtrusive" (T10) or *"quickly ask[s] me for approval, even if I'm doing something else"* (T14).

Teachers wanted to be able to review and change the tool's proposed pairings because they wanted to ensure the pairing choices were good ($N = 9$). They thought *"there may be recent social changes that [affect] the effectiveness of the pair the software may not be able to decipher or be aware of yet"* (T11). If the system auto-paired students, teachers still wanted to be able to distinguish auto-paired students from those already approved by teachers ($N = 4$). Teachers suggested that the tool might "*put the pairing in yellow to show that [they] would not have gotten paired if the teacher was not busy*" (T26). Along with the teachers' preference for being able to front-load preparation tasks, teachers said they wanted to be able to set up *restricted pairing* (i.e., "*pairs that should not happen no matter what.*" [T8]) and *pre-approved* pairings ($N = 5$). Teachers also wanted to monitor students' pairing status and collaboration progress ($N = 4$), "*It would help to not only see who students are working with but also what they are working on*" [T5].

Scenario 3 Results. In the third Scenario (S3), the system proposes a student pairing, shows the teacher details of why these students might work together, and asks the teacher to approve or reject the pairing (Fig. 2). S3 was the *most* favorable scenario among teachers ($M = 5.89$, $SD = 1.13$). The survey asked in a seven-point Likert scale question whether the system should *explain* its reasoning behind the suggested pairing. 77% of teachers agreed or strongly agreed, and only one (2%) teacher disagreed to some extent ($M = 6.07, SD = 0.93$). The survey probed further into what information teachers wanted to see *when approving or rejecting a pairing*. Most teachers responded they wanted to see students' math skill mastery (94%), the problem each student is working on (92%), and students' recent errors (88%). They were much less interested in seeing factors such as students' personality (40%), friendship (32%), and gender (27%) when approving proposed pairings.

Qualitative analysis showed that teachers' overall attitudes towards the scenario were overwhelmingly positive. The majority of teachers ($N = 38$) liked the idea that the system would suggest pairs and give reasoning and justification for the suggested pairs. One teacher expressed, *"THIS IS AMAZING!!! If the system shows me reasonings based on evidence on why [I should] pair some students I would consider it. I love this idea!"* (T43). Teachers found explanations (the scenario did not specify the particular type of explanations) to be valuable as it *"might be something the teacher doesn't realize*" (T1) and could "*provide another pair of eyes*" (T12). Teachers liked that the technology might ensure pairing quality and thought it could increase their work efficiency ($N = 10$). They also liked the idea that in this scenario, even though the system would suggest a pairing, teachers would *have full control* to make final decisions, and the ability to approve or deny the pairing ($N = 11$).

Some teachers ($N = 13$) expressed ideas to further improve this scenario, such as the ability to change pairings occasionally. They also wanted the tool to provide an easy way to find an alternative partner for a student ($N = 2$), such as providing *"a list of other students who would be good pairs"* (T3), or *"a button where I can ask the program to suggest another pair in case I don't like the pair that it suggested"* (T9). Teachers also wanted the system to be accessible (e.g., providing both English and Spanish translation), and compatible with their current contexts and practices (e.g., "*…be able to run on tablets, phones and computers* [T43]" ($N = 4$).

Scenario 4 Results. In the fourth scenario (S4), students suggest their preferred partner, the AI reviews the pairing and notifies teachers of potentially risky pairings (i.e., pairings that may not lead to fruitful collaboration). The teacher then approves or rejects risky pairings and the system pairs all students based on the teacher's decisions. This scenario has the second-highest likeliness of use among teachers ($M = 5.33$, $SD = 1.41$). One of the survey questions asked teachers whether they thought the AI should *notify* them when student pairings are potentially risky and ask them to decide. Most teachers (70%) agreed or strongly agreed ($M = 5.72$, $SD = 1.19$). The survey further asked how teachers wanted students to be notified when they (teachers) rejected student-proposed risky pairings. Some teachers (28%) wanted students to simply be paired with a different partner without further explanation. A similar number of teachers wanted students to be told that *teachers and AI* together rejected the pairing (30%), or the *teachers* rejected their pairings (20%). Some teachers (15%) want students to be told the *AI* rejected their pairing, even though it was, in fact, the teachers who would do so, showing some of them may prefer students to "blame" the AI instead of teachers for not being paired with their preferred partner.

Qualitative analysis showed that teachers' generally viewed the technology used in this scenario in a positive light ($N = 37$). They especially liked that this technology valued students' voice in proposing peer tutors ($N = 16$) and that the systems could act as a safety net and detect if the student-proposed pairings are risky ($N = 13$). Similar to S3, teachers liked the fact that they can make final decisions to approve and reject pairings ($N = 8$). Although teachers thought students' voices were essential ($N = 16$), some were concerned that students' pairing decisions may not be ideal for their learning ($N = 13$). Moreover, many teachers were concerned that rejecting student-proposed pairs may *"affect the relationship between teachers and students and cause unnecessary trouble"* (T53). For example, one teacher thought that *"students should not be able to request in the program. This can lead to many problems in a middle school classroom. Misuse, hurt feelings, etc."* (T1). Teachers also wanted to see the pairing history and results from analysis ($N = 9$; e.g., "*It is hoped that the history of student matchmaking can be added*" [T46]).

Scenario 5 Results. In the fifth scenario (S5), the teachers pair students to work collaboratively, and the AI system reviews teacher-proposed pairings. The AI changes the teachers' proposed pairing when it detects a risky pairing without notifying teachers. S5 was the *least* favorable and most controversial scenario, having both the lowest mean likeliness of use and the highest SD ($M = 4.00$, $SD = 1.76$). To separate design elements teachers like and dislike, the survey then asked teachers on a 7-point Likert scale if the AI should *review* their proposed student pairings. To this, 47% teachers agreed or strongly agreed ($M = 4.87$, $SD = 1.64$). However, when asked whether the system should *override* teachers' proposed pairing if it determined a pair was potentially risky, only 15% agreed or strongly agreed, and 44% of teachers disagreed or strongly disagreed ($M = 3.39$, $SD = 1.86$).

Qualitative analysis showed that teachers' preferences regarding the technology used in the scenario were very divergent. Some teachers liked the fact that the technology could serve as an extra pair of eyes and help them make reasonable pairings ($N = 7$). About one-fifth of teachers expressed that a system that changes risky pairs for teachers can increase teachers' efficiency ($N = 10$). However, the majority of teachers indicated they would decline to use this technology design in their classroom because it would give them too little control ($N = 53$). Specifically, teachers want to have the final say and ultimate control over student pairing; they think they know their students the best and trust their judgment more than the system's ($N = 10$). One teacher explained, "*As a teacher, I will decide what to do and not to do. System modifying without notifying is not the service I am seeking for"* (T5). Instead, teachers wanted the system to *notify* them when it would change their proposed pairings, and allow them to override the system's decisions ($N = 21$).

3.2 Other Results on Teachers' Preferences on Hybrid Control in Pairing

We report how teachers want to share the control of *proposing* and *deciding* with AI and students (RQ1), as stated in the *general preferences* section of the survey.

Regarding ***who should propose pairings***, most teachers (90%) think *teachers* should be involved. Interestingly, more teachers thought the *system* (76%) should be involved than *students* should be involved (53%). In addition, 80% of teachers thought suggesting or proposing student pairing should be *shared*. Among them, the two most popular co-orchestration choices were sharing control *between teacher, student, and system* (41%), or *between teacher and system* (30%). Regarding ***who should make the final decisions*** about pairings, 95% of the teachers thought *teachers* should be the ones to do so. Only two teachers (4%) thought students should make the final decision about student pairing, and only one thought the system should.

4 Discussion

4.1 Insights Related to the Research Questions

We discuss the roles teachers think each agent (i.e., teacher, AI system, students) should or should not take when co-orchestrating dynamic pairing (RQ1), and their preferences regarding dynamic pairing criteria (RQ2).

Teachers' Role. Across all scenarios, teachers prefer to *prepare for student pairing before class, make final pairing decisions,* and *customize the orchestration tool.* Firstly, teachers want to contribute knowledge to help pair up students and to front-load such preparation work prior to class. They want to set pairing restrictions (i.e., identify students who should not be teamed up) and pre-approve pairings. However, many teachers, especially those who teach large classes, do not want to enter information into the system to help the AI pair students, as it may have privacy concerns, or maybe too time-consuming. Given the survey question did not specify either the type or the amount of student information teachers would need to enter, it seems possible that many teachers might be willing to enter a small amount of information, in line with Olsen's finding from co-design studies with teachers [19]. Thus a simple and time-efficient design that allows teachers to set restricted and pre-approved pairings may be needed. Secondly, teachers want to be able to review and modify student pairings proposed by other agents; they strongly want to have the final say. Thirdly, teachers want to be able to customize and configure the co-orchestration tool (e.g., pairing criteria, frequency of changing pairs), to fit their classroom context.

AI Systems' Role. Teachers like the AI to propose personalized pairings, explain the reasoning behind proposed pairings, help evaluate proposed pairings, and lessen teachers' orchestration load. They do not want the AI to make final pairing decisions. Teachers like for the system to propose pairings that are personalized to each student's characteristics. The top three factors that teachers think the system should consider in dynamic pairing are students' knowledge mastery, overall error rate, and past pairing partners. Teachers consider students' friendships, relationships, and gender to be less important in pairing. Most teachers want to use different criteria for different students, suggesting students' characteristics may be weighed in the pairing decision in a manner that varies by situation.

Additionally, teachers want to see reasoning and explanations for why the AI proposes to pair two students or considers a pair to be risky. Teachers want the AI system to act as an extra pair of eyes to review and evaluate proposed pairings, no matter if they are student-proposed or teacher-proposed. Furthermore, teachers want the AI system to lessen their orchestration load, increase work flexibility (e.g., by allowing them to front-load pairing preparation tasks), and minimize distractions while working with students. Interestingly, while teachers want to reduce students' time waiting for their help, they do not all agree that the AI should auto-pair students even when they are busy. Though teachers like being assisted by the AI, most teachers rejected the idea for the AI to pair students without teachers' review or approval.

Students' Role. Teachers see value in allowing students to have a say in the pairing process and provide feedback on pairing, as "*students are more likely to be productive if they are given the opportunity to have a sense of ownership in their partnership*" (T23). One teacher mentioned it may be helpful to notify teachers if "*one partner was unwilling to work [with another student]*" (T41). Compared to teachers' and AI's roles, teachers made fewer comments about the role that students should have, which may be explained by the fact that we have more scenarios describing teacher-AI co-orchestration than student-AI or teacher-student-AI.

4.2 Design Challenges and Directions

Design Challenges. Our findings uncover several *design challenges*. First, teachers' wish to reduce students' waiting time and avoid being the bottleneck blocking students' progress suggests that some degree of system autonomy may be needed. The design challenge is to do so without letting teachers feel their authority in class is threatened. Secondly, there may be a tension between teachers' desired awareness and control, and their desires to avoid interference and distraction when working with students. Relatedly, since teachers often need to make fast decisions when teaching and managing classroom activities, there may be limited time for them to "consume and digest" explanations systems give. It may be necessary to investigate what information will be most helpful to teachers in these explanations and create a design that provides interpretable explanations at a glance. These explanations should take teachers minimal time to read and interpret, *and* still give teachers enough evidence to support their decisions (e.g., approving or rejecting pairings). Finally, it may be hard for a co-orchestration tool to allow students to have a say in pairing decisions and ensure teachers have the final say while avoiding harming teacher-student relationships if students' proposals are rejected.

Design Directions. In connection with the three tasks in dynamic pairing - *propose, evaluate and decide* pairing (Fig. 1, right), our study three design directions, including having the AI system *be a teacher's helper that explains its reasoning (Propose), be an extra pair of eyes (Evaluate),* and *notify instead of deciding (Decide).*

Firstly, it may be fruitful for the AI system to monitor the class and help teachers to prioritize their attention to those who need it the most. For example, if the AI can keep track of students' working progress and pairing needs, and display the class' status to the teacher in *"a queue with time of any students not already being paired*" (T8), it may help teachers to prioritize so students get help sooner than when no AI-monitor is involved. A promising direction is for the AI helper to explain its reasoning when proposing or suggesting certain educational decisions.

Secondly, while teachers want to have the final say in educational decisions (e.g., student pairing), they want the AI to provide an extra pair of eyes, capture things that may escape teachers' consideration, and augment teachers' memory as a form of "distributed cognition." For example, given an AI can keep track of student pairing history, it may suggest to teachers whether or not two students should be paired again. The AI can base its suggestions on the total number of times students have been paired and perhaps their past collaboration quality, information that teachers may not readily have.

Lastly, our study suggests that it may be fruitful to use the AI systems to *notify* teachers of worrying classroom activities that are worth their attention (e.g., risky pairings or unproductive students), and ask them to make final decisions. This may prove much more preferable for teachers than if the AI makes decisions, as it ensures teacher awareness and control. More iterative work is needed to design the orchestration tool, attending to both teachers' preferences and practical feasibility of different pairing policies [20] for pairing students in a given classroom.

4.3 Implications and Outlook

Firstly, this study confirmed, with larger sample size, findings from prior design research and classroom studies [3, 4, 11, 12] that teachers want to: 1) have the final say over pairing students, 2) enable students to get help from other sources when teachers are busy, 3) flexibly front-load preparation tasks for collaborative activities *before class*, and 4) minimize interruptions and orchestration load *during class*. Secondly, we uncover new insights into teachers' needs regarding how to share aspects of dynamic pairing with the AI and with students. The insights include 1) teachers want to see brief reasoning given by the system in at least two situations: when the AI system proposes a pairing, and when it detects a risky pairing. 2) Although teachers want to contribute to student pairings, they are concerned that entering student information into the system may be time-consuming. Thirdly, this study explored how teachers' preferences of orchestration control may depend on certain *dynamic contexts,* including when teachers are busy, when risky pairings occur, when the teacher (S5), system (S2,3), student (S4) proposed pairings.

Future Work. Future work may explore 1) how the dynamic pairing criteria can be personalized based on factors such as class context, students' characteristics, or teachers' demographics and expertise, 2) how co-orchestration tools may afford customizability and adaptability in areas *where teachers have varied opinions*, e.g., whether and how students should be "allowed" to propose pairings, and 3) whether teachers' opinions would change under the time pressure when using the tools in the classroom (e.g., whether and how teachers still wish to review explanations from AI).

5 Conclusion

As researchers start envisioning more sophisticated and personalized interactions in future smart classrooms, fluid social transition become an interesting issue to study [20]. We contribute to the literature of orchestration for fluid social transitions and dynamic pairing. Based on results of the user-centered, mixed-method research through surveying 54 math teachers, the current study extends and complements prior research in human-AI co-orchestration by validating teachers' preferences with a larger sample size, revealing new, nuanced, and context-dependent needs and preferences, and proposing design implications. As the community increasingly adopts a co-orchestration lens to leverage human and AI's complementary strength to achieve synergy, we are hopeful that teacher needs validated and uncovered in this study can help inform future design and research of tools.

Acknowledgement. This work was supported in part by Grant #1822861 from the National Science Foundation (NSF). Any opinions presented in this article are those of the authors and do not represent the views of the NSF. We thank all participating teachers and the anonymous reviewers.

References

1. Hämäläinen, R., Vähäsantanen, K.: Theoretical and pedagogical perspectives on orchestrating creativity and collaborative learning. Educ. Res. Rev. **6**, 169–184 (2011). https://doi.org/10.1016/j.edurev.2011.08.001
2. van Leeuwen, A., Rummel, N.: Orchestration tools to support the teacher during student collaboration: a review. Unterrichtswissenschaft **47**(2), 143–158 (2019). https://doi.org/10.1007/s42010-019-00052-9

3. Olsen, J.K., Rummel, N., Aleven, V.: Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom. Int. J. Artif. Intell. Educ. **31**(1), 24–56 (2020). https://doi.org/10.1007/s40593-020-00228-w
4. Santos, L.P.P.: Supporting orchestration of blended CSCL scenarios in distributed learning environments (2012)
5. Holstein, K., Olsen, J.K.: Human–AI co-orchestration: the role of artificial intelligence in orchestration. In: Handbook of Artificial Intelligence in Education. Edward Elgar Publishing (2021)
6. Tissenbaum, M., Slotta, J.: Supporting classroom orchestration with real-time feedback: a role for teacher dashboards and real-time agents. Int. J. Comput.-Support. Collab. Learn. **14**(3), 325–351 (2019). https://doi.org/10.1007/s11412-019-09306-1
7. Sharples, M.: Shared orchestration within and beyond the classroom. Comput. Educ. **69**, 504–506 (2013). https://doi.org/10.1016/j.compedu.2013.04.014
8. Holstein, K., McLaren, B.M., Aleven, V.: Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 257–266. Association for Computing Machinery, New York (2017)
9. Alavi, H.S., Dillenbourg, P., Kaplan, F.: Distributed awareness for class orchestration. In: Cress, U., Dimitrova, V., Specht, M. (eds.) Learning in the Synergy of Multiple Disciplines, pp. 211–225. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04636-0_21
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 Trans. Learn. Technol. **8**, 187–200 (2015). https://doi.org/10.1109/TLT.2014.2365027
11. Echeverria, V., Holstein, K., Huang, J., Sewall, J., Rummel, N., Aleven, V.: Exploring human–AI control over dynamic transitions between individual and collaborative learning. In: Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M. (eds.) EC-TEL 2020. LNCS, vol. 12315, pp. 230–243. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57717-9_17
12. Holstein, K., McLaren, B.M., Aleven, V.: Designing for complementarity: teacher and student needs for orchestration support in AI-enhanced classrooms. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) AIED 2019. LNCS (LNAI), vol. 11625, pp. 157–171. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23204-7_14
13. Holstein, K., McLaren, B., Aleven, V.: Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In: Rosé, C.P., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 154–168. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_12
14. Davidoff, S., Lee, M.K., Dey, A.K., Zimmerman, J.: Rapidly exploring application design through speed dating. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 429–446. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74853-3_25
15. Holstein, K., Aleven, V., Rummel, N.: A conceptual framework for human–AI hybrid adaptivity in education. In: Bittencourt, I.I., Cukurova, M., Muldner, K., Luckin, R., Millán, E. (eds.) AIED 2020. LNCS (LNAI), vol. 12163, pp. 240–254. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52237-7_20
16. van Leeuwen, A., et al.: Orchestration tools for teachers in the context of individual and collaborative learning: what information do teachers need and what do they do with it? In: Symposium at ICLS (2018)
17. Sankaranarayanan, S., et al.: Designing for learning during collaborative projects online: tools and takeaways. Inf. Learn. Sci. **121**, 569–577 (2020). https://doi.org/10.1108/ILS-04-2020-0095

18. Miles, M.B., Huberman, A.M., Saldana, J.: Qualitative Data Analysis: A Methods Sourcebook. Sage, Thousand Oaks (2013)
19. Olsen, J.: Orchestrating combined collaborative and individual learning in the classroom. https://kilthub.cmu.edu/ndownloader/files/12255413
20. Yang, K.B., et al.: Exploring policies for dynamically teaming up students through log data simulation. In: Educational Data Mining (2021)