Designing Hybrid Human-AI Orchestration Tools for Individual and Collaborative activities: A Technology Probe Study

Vanessa Echeverria, Kexin Yang, LuEttaMae Lawrence, Nikol Rummel, Vincent Aleven

***Abstract*—Combining individual and collaborative learning is common, but dynamic combinations (which happen as-the-need arises, rather than in pre-planned ways, and may happen on an individual basis) are rare. This work reports findings from a technology probe study exploring alternative designs for classroom co-orchestration support for dynamically transitioning between individual and collaborative learning. The study involved 1) a technology-probe classroom study in an authentic, AI-supported classroom to understand teachers' and students' needs for co-orchestration support over dynamic transitions; and 2) workshops and interviews with students and teachers to get informed feedback about their lived experiences. 118 students and three teachers from a middle school in the US experienced a pairing policy – student, teacher and, AI-controlled pairing policy – (i.e., identifying students needing help and potential helpers) for switching from individual to a peer tutoring activity. This work aims to answer the following questions: 1) How did students and teachers react to these pairing policies?; and 2) What are students' and teachers' desires for sharing control over the orchestration of dynamic transitions? Findings suggest the need for a form of hybrid control between students, teachers, and AI systems over transitions, as well as for adaptivity and adaptability for different classroom characteristics, teachers, and students' prior knowledge.**

***Index Terms*—Adaptive and Intelligent Educational Systems, orchestration tools, human-AI orchestration, hybrid human-AI tools, individual learning, collaborative learning.**

I. Introduction

COmbining individual and collaborative learning is common, but dynamic combinations (which happen as-the-need arises rather than in pre-planned ways and may happen on an individual basis) are rare [1]. For instance, many widely used collaborative learning instructional methods, such as the Think-Pair-Share [2] or Jigsaw [3] methods, in fact, use individual phases at some point in the activity to promote a more productive collaboration. In addition, individual and collaborative learning modes may have complementary strengths for supporting learning efficiently [4]. 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 [4]. 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 the 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 pre-planned but are instead determined opportunistically based on unfolding learning situations – whether by an instructor or by educational software.

Manuscript received xxxxx; revised xxx

Orchestrating the dynamic switching in classrooms has been recognized as a major challenge in teaching practice [5], [6]. Although it might seem that orchestration tools designed specifically with this goal in mind could be extremely helpful for teachers, little, if any, past work has focused on designing such tools. Prior research has focused on designing tools for supporting teachers in orchestrating either individual (e.g., [1]) or collaborative learning (e.g., [7]) scenarios, or individual and collaborative learning phases on CSCL scripts (e.g., [8]). 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 [9].

In AI-supported classrooms, such as those using intelligent tutoring systems (ITSs), each student progresses along an individualized learning trajectory, determined by the AI according to a student's needs. During class, teachers may provide additional one-on-one guidance, and help to co-orchestrate the flow of activities in the classroom alongside the AI software. For instance, prior work suggests that during AI-supported class sessions, teachers will sometimes orchestrate transitions on the fly, between individual and collaborative learning (e.g., by pairing one student to tutor another who may currently be struggling) [10], [11]– although they desire greater support from the AI in doing so. Meanwhile, other works have found that students desire some agency over these decisions as well and reject the idea of either teachers or AI systems having full control [10], [12], [13]. In short, these prior works suggest diverse perspectives for providing agency to the different actors, i.e., students, teachers and the AI system, during the classroom orchestration, Still, many open questions remain regarding how best to distribute the task of co-orchestrating dynamic transitions between students, teachers, and AI systems.

In this work, we present a technology-probe study for eliciting design features of a pairing tool and report design challenges and opportunities for human–AI control over dynamic transitions in the classroom. These dynamic transitions (i.e., pairing opportunities) are supported by the pairing tool allowing students, teachers, or the AI system to identify students needing help and potential helpers who might take

0000–0000/00$00.00 © 2023 IEEE

advantage of the benefits of a peer tutoring activity. Thus, three pairing policies for transitioning from individual to a peer tutoring activity were defined: student, teacher and AI-controlled pairing policies. We aim to address the following questions: *1) How did students and teachers react to each of these three pairing policies*; and *2) What are students' and teachers' desires for sharing control over the orchestration of dynamic transitions?*.

II. Prior Work

As pointed out by Dillenbourg and colleagues [1], current CSCL scenarios not only relate to a collaborative activity but also integrate individual (e.g., reading) and class activities (e.g., lectures). In practice, teachers move from individual to collaborative activities following instructional methods, such as Think-Pair-Share [2] or Jigsaw [3], or opportunistically, by matching students struggling with more proficient ones to ease the workload of attending students needing prompt help [14]. In a recent review, authors stated that the orchestration of these transitions does not happen timely, according to classroom needs [15]. In this sense, there has been a surge of interest in designing and implementing orchestration tools in diverse classroom contexts and activities that supports the management and monitoring of diverse learning activities.

A. Classroom Orchestration Tools to Support Dynamic Transitions

Classroom orchestration refers to the management of multiple learning activities and posits the teacher as the conductor of the learning process [5]. Technology can support classroom orchestration by capturing students learning behaviors and informing the teacher about these. The teacher can make informed decisions on the fly and adapt the classroom conditions or instructions in response to the current students' needs [5]. Orchestration tools, then, support teachers in managing and monitoring individual, collaborative or class activities, ideally to lower the teacher load.

Orchestration tools have been developed to support individual activities at the class level, for example, orchestration tools for supporting teachers' awareness using real-time analytics. For example, Holstein and colleagues [16] designed and developed a mixed-reality orchestration tool called Lumilo, which detects students' status and behaviors (e.g., off-task, struggling) while practicing equations in an ITS. Lumilo allowed teachers to quickly observe and pay attention to those needing more help.

Prior works have investigated how orchestration tools allow the adequate supervision of small group activities and improve teachers' awareness of the tasks and interactions happening during these activities (e.g., for a review, see [15]). These orchestration tools may provide different support modes to teachers, such as mirroring, alerting, or guiding modes [15]. *Mirroring* dashboards display basic students' actions and behaviors without providing any interpretation to the teacher. This information often shows students' progress when using a tool (e.g., in an ITS, this could be the number of problems solved, their current level of knowledge, among others) [17]. *Alerting* mode in a dashboard provides information to the teacher about important events needing attention [7]. For example, a dashboard could show real-time analytics about student's learning status based on indicators that capture students' behavior with the ITS (e.g., when a student is idle or is making many errors) so the teacher can make an informed decision and act promptly [17], [18]. Finally, *advising* dashboards provide suggestions about the current student's status and possible ways to act. In an ITS, this advising information could be provided as suggestions to pair up students who are not progressing in their math skills with students who already mastered a specific set of skills [17].

Few studies have investigated the design and implementation of dynamic orchestration tools for transitioning from individual to group activities to a lesser extent. For example, [14] reported a set of tools to assist teachers in individual activities by alerting them which students need immediate help, finding students' status (e.g., task completion, struggling, disengagement when using an intelligent tutoring system), and in group activities, by showing pairing suggestions for supporting group discussions. Similarly, the work presented in [8] evaluated an orchestration tool to support dynamic transitions in a collaborative inquiry activity. The tool sends the teacher notifications about the state of the class while interacting in an equipped smart environment (i.e., tablets, multitouch tables, projectors), either in the individual or collaborative phase of the activity. It also matches students to work in groups and monitors their progress. While these few works illustrate how orchestration tools could timely support teachers' classroom management when transitioning from individual to collaborative learning, students switch simultaneously from one instruction to another. By contrast, our study aims to investigate the design needs for orchestration tools to support dynamic transitions as-the-need arises.

Other works have pointed to the need for shared orchestration (or co-orchestration) for individual and collaborative activities, given the complex and dynamic scenario. In a recent work by Olsen and colleagues [13], authors reported a set of design desires for orchestrating dynamic transitions but focusing on the shared roles and responsibilities of intelligent systems and humans (i.e., teachers and students). Also, a study reported by Holstein and colleagues [10] found that one of the most preferred concepts among students is the ability to have a student-system shared control over the selection of peer tutors. The authors also reported a contrasting view from teachers, who preferred to work with the AI system to match peer tutors and tutees. Our work extends these current studies by eliciting the need for shared control between students, teachers and the AI-system from teachers and students.

B. Group Formation for Effective Dynamic Transitions

Orchestrating dynamic transitions involves a group formation process in which pairs of students are selected and are expected to collaborate effectively. This issue has been studied extensively in Computer-Supported Collaborative Learning (CSCL) research (see reviews in [19], [20]). Several considerations and techniques when forming groups lead to effective collaboration. In this section, we review group formation in terms of 1) students' characteristics, 2) group type and 3) participants involved in the group formation, which are relevant to our context.

Prior works have reported diverse *students' characteristics* to find compatible members that lead to a positive learning outcome [20]. For example, gender or task proficiency has been used as input data to find suitable members and conform groups. In our work, we consider students' skill levels as a key characteristic to determine who may need help or may be a good candidate for becoming a tutor.

Concerning the *group type*, studies in CSCL have mainly reported heterogeneous and homogeneous group types (see [19], [20] for a complete overview of group formation types). For instance, in homogenous group formation, participants with similar characteristics are considered potential group members. In practice, students involved in the selection process tend to create homogeneous groups (e.g., based on their affinity and similar knowledge levels). In contrast, in heterogeneous group formation, participants with diverse skills are selected to support each other learning. In practice, teachers often tend to create heterogeneous groups based on their beliefs about students' prior knowledge. AI systems could intervene in the group formation process and support creating heterogeneous or homogeneous groups by accessing students' level skills and determining which students could take advantage of collaboration. In this study, we explore homogeneous pairings (those initiated by the student) and heterogeneous pairings (those initiated by the teacher or the AI system).

Finally, some researchers have reported how the *involvement of different participants* – students, teachers and the AI system –in the group formation process can affect the learning experience [21]. In some cases, students can be asked to specify their preferences for collaboration (e.g., based on affinity or social relationships). This is not always ideal, as students may not know their learning needs or knowledge levels. In other cases, the teacher have full agency over the whole group formation process by initiating and identifying peers based on their experiences or beliefs. However, this can be time-consuming or difficult when the teacher does have a prior conception of students' knowledge. Intelligent support can be obtained from an *AI system*. The AI system can automatically form groups by considering students' and teachers' inputs. Nevertheless, there is still scarce evidence of the benefits and desires for fully automating this process. This study examines the student, teacher, and AI system participation in the group formation process.

III. Context

A. AI-Based tutoring systems for individual and collaborative learning

As mentioned above, many instructional methods include individual and collaborative phases to improve students' learning gains [2], [3], [8]. However, these methods usually unfold in a pre-planned, synchronized manner, where the teacher sets the time limit for each phase and students switch simultaneously. In the current work, we aim to support fluid dynamic transitions, where the teacher monitors students' status and, according to students' needs, teachers or the AI system could decide to initiate a peer tutoring activity. Real-time AI systems could support these transitions by capturing students' interaction logs, and the AI system can group students according to teachers' inputs or students' performance [8], [22]. In this way, AI systems may support teachers by adapting the instructional method according to the classroom dynamics and potentially ensuring productive outcomes [23].

In our study, students dynamically transitioned between two AI-based tutoring systems, namely, Lynnette (see Fig. 1), which supports individual problem-solving practice, and APTA 2.0 (see Fig. 2), which supports mutual peer tutoring, a form of collaborative learning. Both systems support practice in linear equation solving for middle school students.

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 [24] within the CTAT/Tutorshop architecture [25].

To support collaborative learning, we implemented a new version of **APTA** (Adaptive Peer Tutoring Assistant), developed originally by Walker, Rummel, and Koedinger [26]. This system adaptively coaches one student (the "peer tutor") in tutoring another student (the "tutee") with advice about both tutoring and mathematics. It uses two rule-based cognitive models, one that captures peer tutoring strategies and one that captures equation-solving knowledge (the latter is shared with Lynnette.)

APTA 2.0 supports two different interfaces, one where the tutee (the student being helped) solves linear equations (Fig. 2 - top) and another through which the student in the peer tutor role monitors the tutee's work and provides guidance (Fig. 2 - 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 the 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. 2 - bottom).

APTA 2.0 gives the peer tutor feedback on whether their marking of the tutee's steps is correct. To this end, APTA 2.0 connects with Lynnette, which compares the tutee's input to possible correct solutions generated by its cognitive model of equation solving. For example, in Fig. 2 we can observe that if the tutor marks the tutee's step as incorrect, the × button is highlighted in green, indicating that the tutor's grading is correct. At the same time, the tutee's step is highlighted in red, indicating that the solving step for the equation is wrong.

APTA 2.0 also presents a chat module where the tutee and the tutor can communicate during the assignment. For example, the tutee can ask for help, and the tutor can give hints on the current step (Fig. 2, chat component). Chat messages are classified as help types (e.g., next-step help, previous-step help, both and not help) [26]. The classification result is then used to feed the coaching model and provide adaptive advice

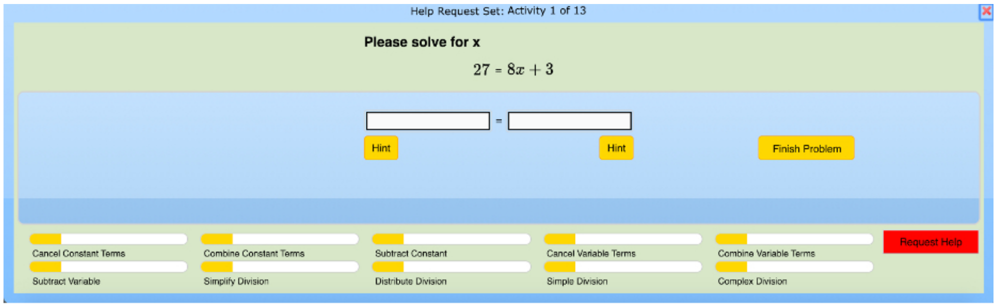

Fig. 1. Lynnette interface for individual assignments.

to the tutor, as depicted in Fig. 2. APTA 2.0 assesses students' collaboration skills using its model of collaboration with a variant of Bayesian Knowledge Tracing.

APTA 2.0 also has a teacher's interface, where a teacher can log in, access her classes, and add a roster. She can also configure problem sets and skills, check students' progress and skills in real-time, and manually assign pairs to initiate a peer tutoring activity. For example, Fig. 3 (A) depicts the *Reports view*, where the teacher can access the list of students and observe their progress on the problem sets they have been assigned. This view also allows filtering of the information based on students skills and performance based on the assignments students are working on. The legend of the student's progress can be found at the bottom of the view. All these reports can be accessed in real-time. Fig. 3 (B) illustrates the *Assignment view*, where the teacher can see all students who are working on a specific assignment. The teacher can also choose students to work on this assignment and initiate a peer tutoring activity. Fig. 3 (B) shows an example of a peer assignment called *SubstractConstant*. Two students have been assigned to the *SubstractConstant* assignment and have already completed the task.

While Lynnette and APTA have been used in prior studies separately and have each separately shown improved learning gains [27]–[29], the current study is the first attempt to combine both AI systems for dynamically switching between individual and collaborative learning activities.

B. Dynamically transitioning from individual to peer tutoring activities

In the current study, we support fluid dynamic transitions, where the teacher monitors individual students' status and, according to students' needs, they can work on tutoring activities or keep working individually with the help of an intelligent system (i.e., Lynnette). Fig. 4 depicts the orchestration tasks for dynamically transitioning between individual and peer tutoring activities. In principle, any of the participants (e.g., students, teachers and/or the AI system) could have control over any of these tasks - although, as mentioned, a key question is how control should be shared or divided - and how the orchestration tool should be designed to support the desired sharing or division of control. It is worth noting that there are more cases to be considered in these three main tasks (described below) for transitioning from individual to peer tutoring activities and that the examples provided here help illustrate the type of control and decisions that the participants may be involved in.

- **Task 1: Identifying students needing help and potential partners:** In prior works, researchers have explored how technology can support group formation to maximize individuals' and groups' outcomes [30]. This is usually done by gathering students' characteristics, so the system or teacher can make informed decisions for selecting potential partners to conform to a group based on similar (homogeneous group formation) or complementary needs (heterogeneous group formation) (e.g., [8]). The student, the teacher, or the AI system could control this task. For students, this task could mean asking for help if they struggle with the equation-solving exercise. Teachers could identify struggling students by considering their prior knowledge or by observing students' behavior during the individual activity and then detect potential partners that can act as mentors. As for the AI system, it could detect struggling students (c.f. [12]) working in an individual activity and could provide a potential match with a student that could act as a tutor (i.e., a student that has already mastered a specific skill and have good aptitudes to be a helper).
- **Task 2: Negotiating with participants:** Negotiation is a complex process where all participants could be involved in the decision of accepting or rejecting a new task based on their goals, preferences and interests [30]. Classroom dynamics and instructional goals could be considered to delineate which participants should have the power or authority to accept and/or reject tasks. Students could decide if they are willing to participate or not in a new peer tutoring activity. For example, a high-performing student could reject a request made by another student because she may want to finish her work first and does not want to be interrupted. Teachers could decide if they accept or reject pairing suggestions from students or the AI system. For example, following a request by a student to collaborate with another student, the teacher may judge that it would not be beneficial for this student to start a new peer tutoring activity because it could be a distraction. Finally, for the AI system, this negotiation could be related to the co-configuration and optimization of peer tutoring activities. For example, the AI system could suggest that the peer activity initiated

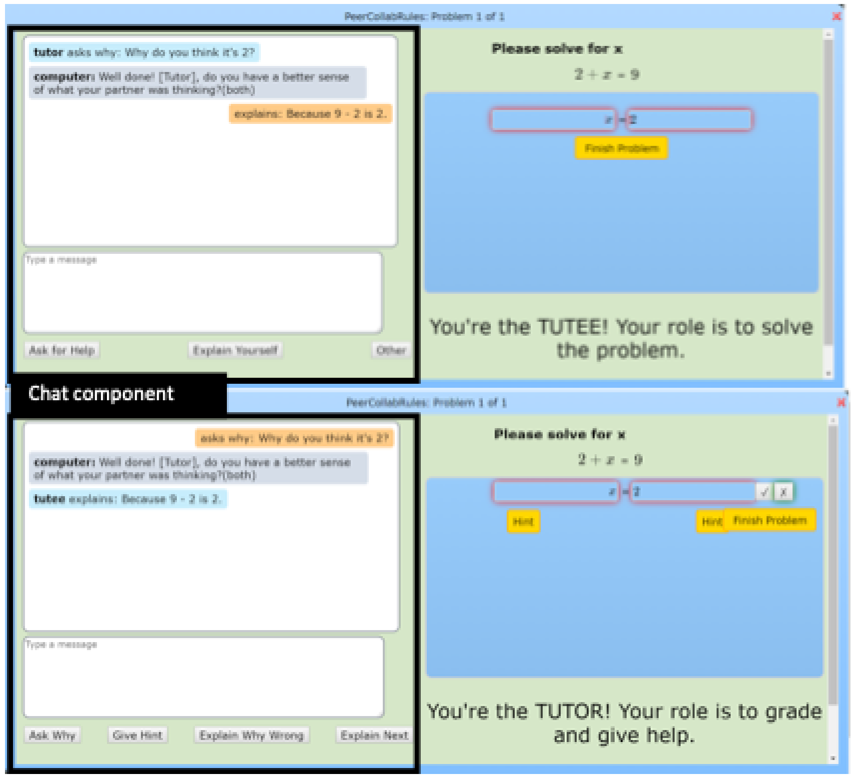

Fig. 2. APTA 2.0 interfaces for students in the tutee (top) and tutor (bottom) role.

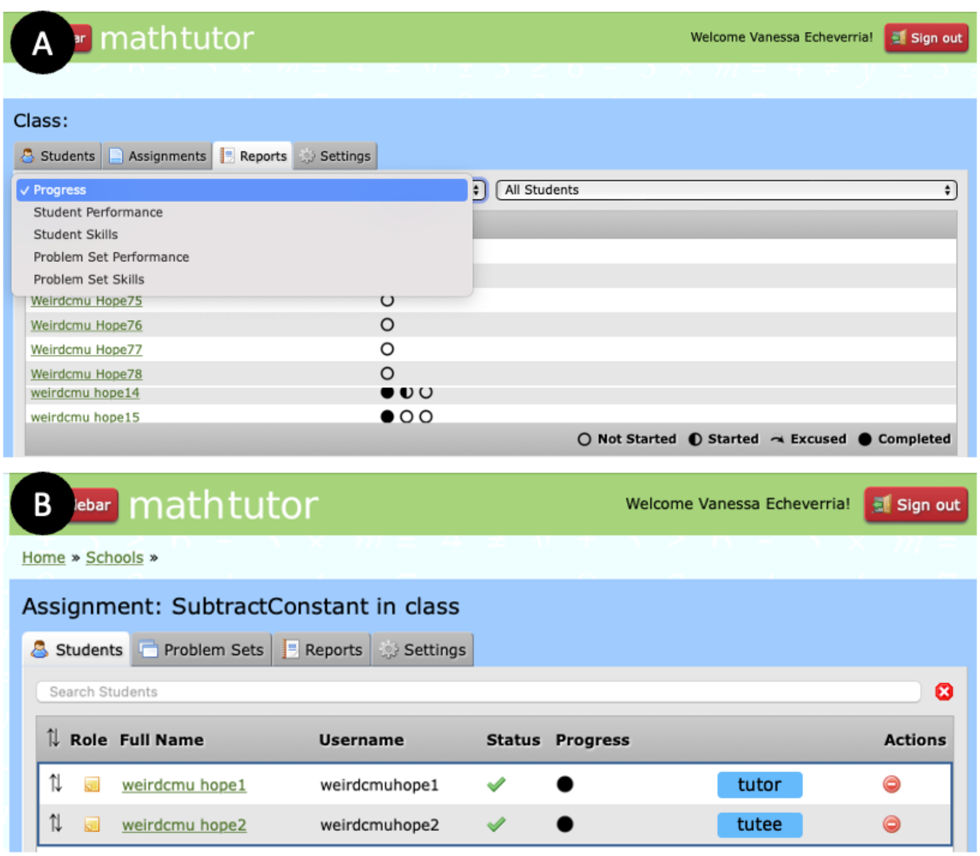

Fig. 3. APTA 2.0 teacher interface. (A) Reports view - where the teacher can access students' progress, performance and skills. (B) Assignment view - where the teacher can choose a pair of students to work in a tutor-tutee activity.

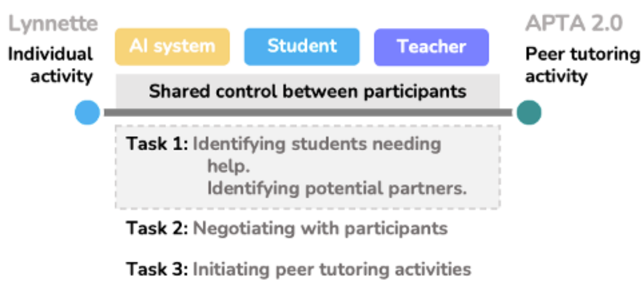

Fig. 4. Representation of the shared control between participants (i.e., AI system, student, teacher) during dynamic transitions.

by the teacher might not be productive and then provide some recommendations for effective pairing.

- **Task 3: Initiating peer tutoring activity:** Finally, participants could have the control to initiate the peer tutoring activity [13], [30], resulting in the instantiation and enactment of this activity. This initialization depends on students' needs (i.e., past knowledge, and current skills) and the decisions taken in the previous tasks. For instance, a student could initiate a peer tutoring activity that has already been negotiated with the teacher or with another student. A teacher could initiate a peer tutoring activity based on prior students' knowledge without requiring students' approval. And finally, the AI system could also initiate a peer tutoring activity without the teacher's or students' input or approval.

When all steps of the dynamic pairing process have been completed successfully, a newly-formed pair of students will be working on a peer tutoring activity using APTA 2.0. The other students will continue their work (either individually with Lynnette or collaboratively with APTA 2.0). The current work focuses on Task 1, namely the **pairing policy**. The other two tasks are being further explored in ongoing research [31].

In the next section, we introduce the methodology used to explore the boundaries for co-orchestrating the pairing policy task between teachers, students, and AI systems.

IV. Methodology

A. Technology Probe Study Design

To explore designs of classroom co-orchestration support for dynamically transitioning between individual and collaborative learning, we conducted a technology probe study in middle school classrooms. Following Hutchinson et al.'s conception of technology probes [32], 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 version 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. The technology probe study was designed to explore alternative designs within a subset of the design space we described, namely, the task of identifying students who might benefit from collaboration and finding potential partners for them.

Based on prior design explorations with students and teachers [10], [13], we defined three pairing policies for dynamically transitioning between individual and collaborative learning. At this point, classroom characteristics (i.e., low-achieving, high-achieving classes) were unknown, as the study intended to understand how dynamic transitions may unfold in authentic classrooms. The steps to initiate and actuate a switch from one learning mode to another for given students were not yet fully automated, as we envision they will be in the future; they were conducted following a Wizard of Oz (WoZ) experiment (i.e., a remote researcher interacted with the tool to initiate a peer tutoring activity). Each class was randomly assigned to one of the three policies described below.

Student-controlled pairing 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 an online request form (i.e., Google Forms). In this policy, a student (tutee) requested help by clicking the "Request Help" red button in the Lynnette interface (see Fig. 1), meaning that the tutee had full control over the selection of tutor candidates and there was no negotiation with the teacher and the AI system. Once the request was delivered, the WOZ AI system (simulated by a remote researcher) initialized a peer tutoring assignment in APTA 2.0 by matching the tutee with the first option listed on the submitted form. If that option was not available (e.g., because the requested partner was working on another peer tutoring assignment), the AI system tried to match the tutee with the second option listed, and so on, until fulfilling the help request.

Teacher-controlled pairing policy. Teachers were encouraged to identify students who were struggling (tutee) and pair them with students who could be of help (tutor). Teachers selected students primarily based on their conception of the student's knowledge and skills. Upon the teacher's request, information about each student's skill mastery in Lynnette was shown to them. The teacher identified a student (tutee) that could potentially benefit from a peer tutoring activity and a partner (tutor) and requested the AI system (simulated by a remote researcher) to pair them up and initialize a new peer tutoring assignment in APTA 2.0.

AI system-controlled pairing policy. Before fully deploying an AI system policy, we used a WoZ AI system pairing policy to explore the configuration and run-time of an intelligent pairing policy in an authentic scenario [1]. Following a technology probe approach, the simulated AI system was constantly monitoring students' skills (from Lynnette) and identifying students who had a lower skill mastery (as candidates for the tutee role). Then, the AI system found a possible candidate with higher skills (tutor), assuming that students with lower skills could benefit from peer tutoring activities. The remote researcher was instructed to identify students who, for one of the ten skills being monitored (e.g.,

[1]A complete exploration of an existing AI ecosystem for pairing students has been reported in [17].

cancel constant terms, as shown at the bottom of Fig. 1), had a probability of knowing – as estimated by Bayesian Knowledge Tracing [33] – below 50% for the tutee candidate and above 75% for the tutor candidate. The remote researcher had access to students' skill mastery from Lynnette's teacher's view. If the tutor was already paired up with another student, then the researcher would choose the next best match, based on any other skill, and ultimately choose a random tutor who was not paired up yet. The AI system initialized the new peer tutoring assignment in APTA 2.0.

B. Participants and Procedure

Three seventh-grade math teachers (1 female, 2 male) from a middle school in a suburban school district in the US were recruited. Teachers were asked to use both AI-based tutoring systems during their normal classroom period. A total of 118 students from six classes (C1 – C6; see Table 1). Students practiced equation solving for further skill development and refinement.

During two regular class periods, each lasting for 45 minutes, students performed a set of tasks using Lynnette and/or APTA 2.0, respectively, for individual and peer tutoring assignments. First, students followed a mini-tutorial on how to use Lynnette and APTA 2.0. Second, students were engaged in individual activities by solving four sets of individual assignments using Lynnette. Third, after 20 minutes of working on individual assignments, a phase started in which some students were paired up dynamically for collaborative learning, using the pairing policy in effect for the given class, as explained above. Peers (tutee and tutor) were asked to stop their individual assignments and work on an assignment with a partner. Fourth, students participated in a discussion workshop led by the teacher to discuss their lived experiences (e.g., *Did you like to be paired with a peer to solve linear equations? Would you prefer to select your 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?*)

TABLE I
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

Afterward, teachers participated in post-hoc 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 interview sessions, each lasting about 30 minutes: one with two teachers (A and B) and the other with one teacher (C). Given that only one of the three teachers experienced all pairing policies, we prepared a set of storyboards[2] representing the three different pairing policies (student-, teacher-, or AI system- controlled) as described above to help the teachers understand the differences between these. Teachers were asked to review these storyboards and explain their preferences on co-orchestration opportunities (e.g., *Who should have control over the pairing policy? Who should accept or reject the initialization of a peer tutoring activity?*).

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. Similar to the approach in [18], the observer took observational notes regarding teacher and students' behaviors using LookWhosTalking[3]. The tool allows for coding live classroom observations, customized with pre-configured categories for teachers and students (e.g., the teacher explaining instructions to the whole classroom, a student talking to another student). These categories were grounded in prior works, such as [18], [34], where they developed a protocol for exploring and understanding teacher and students' interactions. The observer was trained to conduct observations in a pre-training class session. The researcher took notes during students' workshop sessions and teachers' interviews. All logged data generated by Lynette and APTA were collected in the PLC DataShop repository [25] for further analysis.

V. ANALYSES AND RESULTS

To understand design challenges and opportunities for human-AI control over dynamic transitions in the classroom, we aim to answer the following research questions: A) How did students and teachers react to each of these three pairing policies? And B)What are students' and teachers' desires for sharing control over the orchestration of dynamic transitions?

We analyzed quantitative data collected from the Lynnette and APTA 2.0 log data and qualitative data collected from classroom observation notes from six classes, along with notes from post-hoc workshop discussions with students and semi-structured interviews with teachers. Quantitative data were analyzed and summarized to understand individual and peer assignments and to understand and illustrate the variability in orchestration dynamics at the classroom level. . Qualitative data were analyzed following a content analysis procedure [35]. Quotes of interest were selected by two researchers and then summarized in relation to the experience of dynamic switching and in relation to each pairing policy.

A. RQ1: How did students and teachers react to the pairing policies?

We first obtained the number of students who engaged in individual assignments and peer tutoring assignments (excluding the mini-tutorial assignments) for each policy using the data logs from Lynnette and APTA 2.0 (see Table I).Fewer students engaged in peer tutoring assignments under the student-controlled pairing policy than under the teacher-controlled or the AI-controlled pairing policy.

[2]The full protocol can be found at https://vanechev.github.io/cyberlearning/teacher-protocol.html

[3]Available at: https://bitbucket.org/dadamson/lookwhostalking

This article has been accepted for publication in IEEE Transactions on Learning Technologies. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TLT.2023.3248155

TABLE II
TIME SPENT AND ERRORS MADE ON PEER TUTORING ACTIVITIES IN EACH CLASS AND PAIRING POLICY – AVERAGE (Q3 – Q1)

Policy	Class	pairs	time spent (secs) avg ± stdev (Q3-Q1)	average errors made avg ± stdev (Q3-Q1)	Pairs that completed all assignments
Student (n=41)	1	2	96.5 ± 47.6 (113.4-79.7)	2.0 ± 2.8 (3.0-1.0)	2 (100%)
	2	6	188.6 ± 95.5 (227.8-132.7)	7.0 ± 10.9 (22.0-0.0)	5 (83%)
Teacher (n=31)	3	7	168.4 ± 72.0 (222.0-120.7)	11.0 ± 10.0 (21.0-2.5)	4 (57%)
	4	6	**270.0** ± 138.3 (333.4-159.4)	**30.7** ± 27.2 (47.0-10.8)	2 (33%)
AI system (n=46)	5	10	138.8 ± 97.3 (211.0-53.9)	8.0 ± 12.0 (13.0-0.0)	7 (70%)
	6	9	216.0 ± 205.9 (258.2-91.1)	3.6 ± 6.1 (2.0-0.0)	4 (44%)

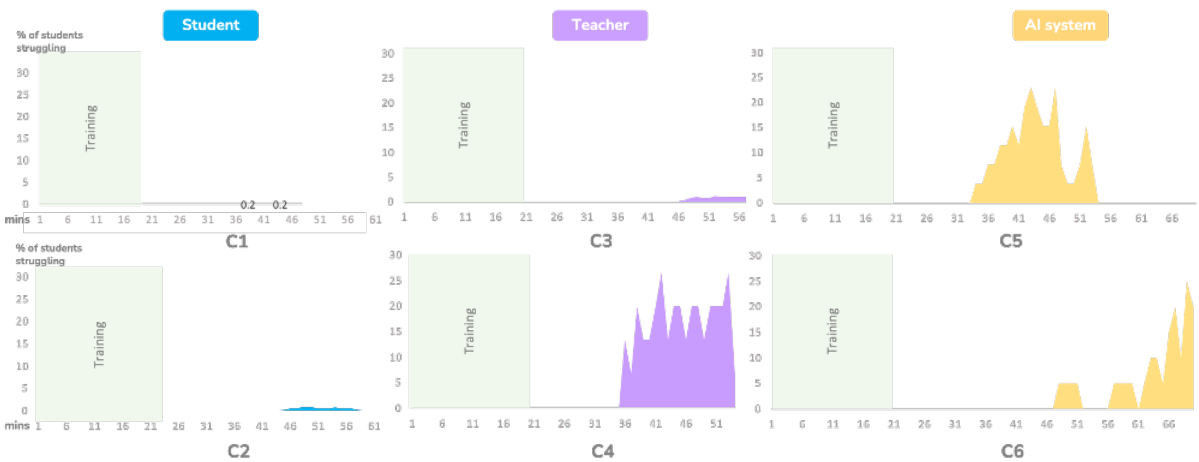

Fig. 5. Percentage of students struggling during the pairing activity on each class and pairing policy.

A summary of the average time spent (in seconds), the average of errors made, and the number of pairs that fully completed the peer tutoring assignments are presented in Table II. From Table II, we can observe that pairs from C1 spent less time (M = 96.5 secs; SD = 47.6); had fewer errors (M = 2 errors; SD = 2.8) compared to the rest of the classes, and all the pairs completed all the assignments. On the other hand, pairs from C4 spent more time (M = 270 secs; SD = 138.3), had a higher average of errors (M = 30.7 errors; SD = 27.2) and 33% of pairs completed all the assignments.

In addition to the summary presented above, we wanted to explore if students who were engaged in a peer tutoring activity had a productive collaboration. Thus, we computed (for each minute of the collaborative episode) whether the tutee was struggling using the same struggle detector as [36], as a proxy of an unproductive collaboration episode (see Fig. 5) Students (tutees) that struggle during peer tutoring activities are not progressing with the help of the student tutor. Next, we summarize the interpretation of our results in terms of the three pairing policies and teachers' and students' responses.

Student-controlled pairing policy: In the *student-controlled* 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 in the two classes that experienced this condition (C1 and C2, see Table I), we analyzed the behavior of these two classes separately from log data, observations, and interviews. Only four students from C1 requested to be engaged in peer tutoring assignments. From classroom observations and the teacher's responses, we found that students in this class were more confident and had more advanced math skills than those in C2 (i.e., *"class 1 is a high-achieving classroom" - Teacher A*). Students in C1 expressed that they did not feel the need to ask for help and initiate a peer assignment. As for students from C2, half of the students worked on a peer tutoring assignment, a low number compared to the other two conditions (e.g., 14 out of 16 students from C3 worked on a peer tutoring activity). From classroom observations and students' responses, we found that not all students were equally motivated to work on collaborative learning activities. For instance, most students from C2 stated that *"they would prefer working alone."* Pairs from C1 and C2 spent a moderate amount of time working on the assignments. C1 had fewer errors (only two pairs were created), and C2 had an average of 7 errors (from the 7 working pairs). Almost all pairs completed all the assignments (C1: 100% and C2: 83%), and their struggling statuses were low (see Fig. 5 - left column).

Concerning students' experience with the pairing policy, most of the students preferred having full agency over choosing their own partners. For instance, all students from C1 (high-achieving classroom as mentioned by the teacher and defined in Section V - B) and some students in C2 (a large classroom, as defined in Section V - B) mentioned they liked being able to choose a classmate and not being paired by the teacher. A student from C1 expressed that he would prefer *"to choose someone [he] 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"*, which may suggest that students would like to have access to their skill mastery levels to choose their best partner. Most students in C2 expressed 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, which may suggest a student-teacher shared control over the pairing activity, i.e., students detecting and signaling (to the teacher) that they are struggling and the teacher choosing the partner.

The teachers' perspectives did not align with students' preferences. Teachers expressed that they prefer unrestricted control over pairing. For example, 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 decision over the pairing is to reduce non-math-related chat interactions on the part of their students. Findings from prior work and analysis of chat interactions from the current study seem to suggest that teachers have reason to be concerned: It is very common for students to use the chat for interchanging off-topic messages.

Teacher-controlled pairing policy: In the teacher pairing policy, the teacher was the instigator of the dynamic pairing. Table I shows that 26 out of 31 (83.9%) of students in this condition engaged in a peer tutoring activity. The teacher could recognize students who might benefit from peer tutoring and pair up students either by herself (7 pairings requested) or by requesting information about students' skill mastery, which was retrieved from Lynnette (6 pairings requested). From Table II, we can observe that C4 spent more time and made more errors compared to other classes in the teacher-controlled pairing policy. Not all pairs completed all the assignments (C3: 57% and C2: 33%). Struggling status in C3 was low and in C4 relatively high (see Fig. 5 – middle column).

Regarding students' experience with the pairing policy, most students from C3 and roughly half of the students from C4 (a small classroom, as defined in Section V - B) positively reacted to letting the teacher make pairing decisions, suggesting that the teacher should have full agency over choosing partners. One student stated: *"[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, similar to students' perspective on the student-controlled pairing policy, some students from C3 and C4 stated that they would like to choose a classmate as tutor candidate *"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 (i.e., student having a full agency to choose a partner). Otherwise, they would prefer to receive help from the teacher rather than a peer. Although most students were positive about the pairing decisions made by the teacher, they also recommended other pairing policies. For example, one student mentioned that she would prefer *"to get a randomized partner because she can get someone new every time"* (i.e., shared student-system control when a student requests a random tutor). 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 [potential tutor] skill" or "a qualification of becoming a tutor,"* and only students who have this skill should be recommended for tutoring other students. As explained above, students did not have access to any up-to-date information about other students' math or peer tutoring skills. These comments from students raise the interesting novel idea of a pairing policy that would consider APTA's assessment of a student's peer tutoring skills. (As mentioned, APTA 2.0 uses its model of tutoring skills to assess individual students' skills in this area.) Therefore, it may suggest a shared student-system control over the pairing activity.

As for teachers' perspectives, they acknowledged that this 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 the teachers' side, as one teacher expressed that *"at some point, matching and monitoring individual and peer tutoring activities would be bothersome."*

AI System-controlled pairing policy: In the AI system policy, 38 out of 41 (82.3%) students were engaged in peer tutoring assignments, similar to the teacher pairing policy. Our records indicate that, of the peer tutoring activities under this pairing policy, 78.9% of them were chosen based on the same skill for the tutee (below 50%) and the tutor (above 75%); 15.8% were chosen based on the next best match by considering any higher skill for the tutor, and 5.23% were initialized by selecting a random 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 with differential mastery of the same skill. From Table I, we can observe that pairs in C6 spent more time but had fewer errors than C5. Almost all pairs in C5 completed all the assignments (70%), but fewer pairs in C6 completed all the assignments (44%). The amount of struggle (as measured by the struggle detector) for C5 was relatively high at the beginning of the peer tutoring activity but decreased towards the end of the activity. In contrast, the amount of struggle for C6 increased towards the end of the activity (see Fig. 5 – right column), suggesting that students in the peer tutoring activity were unable to improve their skills. From classroom observations and teachers' responses, we noticed that C6 was a *"struggling classroom"*. Teachers B and C mentioned that some students had basic or limited knowledge of the given subject matter (linear equation solving). They also expressed that some students had external issues that may hinder their learning.

Regarding students' experience with the pairing policy, most students from C5 (large classroom) expressed no reservations about being paired up by the system, suggesting full agency to the system over the pairing activity. However, most students from C6 (struggling classroom) had a contrasting view, expressing that they would prefer to choose a classmate to become their tutor (i.e., students' full agency over the pairing activity). 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 from previous 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."*

As for teachers' perspectives, they liked the idea of having an AI system pair up students. However, teachers pointed out they should have a certain degree of control over the AI system's decisions, suggesting a shared teacher-system control over the pairing activity. One teacher mentioned that *"she must be able to override the system's matching decisions."* Teachers also suggested that the system should have some constraints for matching students based, for example, on students' characteristics, social dynamics and prior knowledge. For example, teacher C expressed 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 accepting or rejecting a matching suggestion.

B. RQ2: What are students' and teachers' desires for a shared control over dynamic transitions?

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 by the results presented for RQ1. We explain this view according to three differentiated classroom characteristics that emerged from teachers' responses.

In addition, we analyzed how students' degree of struggle and their initial knowledge differentiated among classes and how it might have influenced teachers and students' preferences for control of dynamic transitions. Students' struggle is defined as students making many errors without reaching mastery in skills that they are practicing [16]. *Struggling behaviors* have been linked with students' lack of learning, frustration, and disengagement [37]. We analyzed the ratio of students who were in a struggle status (Ratio of Struggle Status - RSS) for the first 20 minutes of working on an individual assignment using a struggle detector [37]. The detector is implemented in JavaScript, categorizing students as struggling if their number of total correct attempts at steps in a math problem within recent attempts (window size = 10) is below a certain threshold ($n = 3$). Fig. 6a shows the RSS for each class. Similar to the analysis approach presented in [36], the ANOVA results indicated that the classes have different ratios of struggling students (RSS) [$F(5,62) = 1.94$, $p < 0.1$]. Post-hoc Tukey tests showed a significant difference in the RSS between C3 and C1 (diff = 0.09, $p < 0.01$), C3 and C4 (diff = -0.07, $p < 0.05$), C3 and C5 (diff = -0.08, $p < 0.05$).

Students' *initial knowledge* (IK) measures students' pre-mastery skills (i.e., equation-solving skills), indicating students' baseline learning on these skills. For the *initial knowledge analysis*, we calculated the average of IK from all Knowledge Components (KC) involved in each individual assignment. Specifically, for each KC, we used the average mastery level (predicted by BKT) from the first three attempts each student made to represent their IK on this KC. We then took the average IK across all KCs involved for one student to represent their overall IK. Fig. 6b shows the IK averaged across students per class. From the Kruskal-Wallis Test, results indicated a significant effect of classes on IK for the six classes [$F(5, 120) = 17.6$, $p < 0.01$], and the IK for the six classes were not all equal.

From the quantitative analyses and the qualitative data gathered from classroom observations and teachers' interviews, we illustrate three differentiated classroom characteristics and describe the preferences for co-orchestration of dynamic transitions.

High-achieving classrooms. From the teachers' comments, we found that C1 was a high-achieving classroom. Teacher A stated that most of the students from his class were taking advanced math classes. There were no significant differences in the proportion of struggle or in students' initial knowledge. Also, as indicated in the results of RQ1, only four out of 17 students worked on peer tutoring activities (see Table I), in which students expressed they did not feel the need to ask for help. These pairs had fewer errors (see Table II) and did not struggle during the peer assignment activity (see Fig. 5). Concerning the co-orchestration of dynamic transitions, teachers suggested that for high-achieving classes, the control over the pairing policy could be ***shared by students and teachers***, giving partial agency to students by letting them choose a partner to work with, with the teacher having the option to *accept or reject* them. In line with prior findings [10], [38], teachers agreed to give students some control over the pairing.

Struggling classrooms. We found that C6 was a low-achieving classroom from teachers' comments, classroom observations and quantitative analyses. Although there were no significant differences in the struggling and IK analyses, both indicators suggest a higher ratio and variability of struggling students (Fig. 6a) and higher variability of IK, ranging from 0.2 to 0.6 (Fig. 6b). As noted in the previous section, 18 out of 20 students worked in peer tutoring activities (see Table I). Working pairs in C6 had the second highest average of time spent in the peer tutoring activity (avg= 216.0, Table II); only 44% of pairs completed the peer tutoring activity. Furthermore, their teacher (Teacher B) stated that there were several students with math scores below the average and that usually, the classroom dynamics for this class are different from others (e.g., large classes like C2). Concerning the co-orchestration of dynamic transitions, in line with prior findings, teachers indicated that for this class, orchestration could be the ***shared between the AI system and teachers*** [10], [13]. For example, teacher C indicated that she would let the orchestration tool match students according to students' skills if the system is *"able to restrict some matchings"* depending on students'

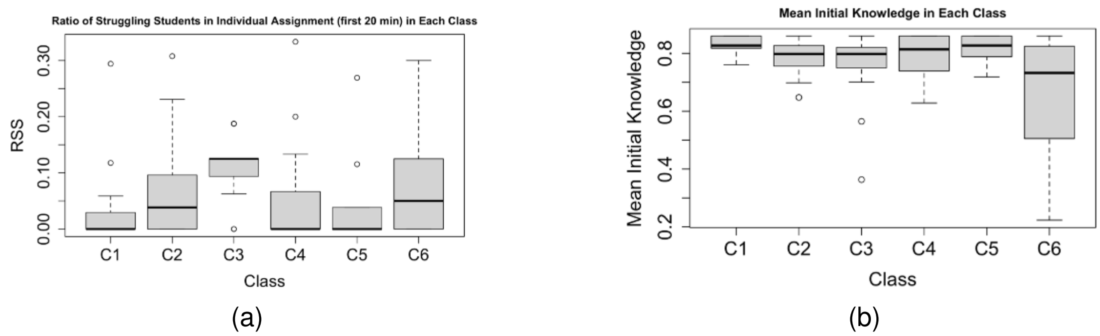

Fig. 6. (a) Ratio of struggling students and (b) mean initial knowledge in each class C(1-6)

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 indicated 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 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. We considered a class to be large if the number of students is above the average class size in the U.S. (n=16.6 for middle schools)[4]. Therefore, we considered C2 and C5 as large classes. C2 and C5 experienced student and AI system-paring policies, respectively. Almost all working pairs in C2 (83%) completed their peer assignments. C2 had a lower ratio of struggling students during their peer assignments (see Fig. 5, left column). In contrast, C5 had the highest average time spent (avg= 270.0, Table II), the highest average of errors made (avg= 30.7, Table II) and a lower percentage (33%) of completed peer assignments. C5 also had a high ratio of struggling students during the peer tutoring activity (see Fig. 5, right column). Regarding the co-orchestration of dynamic transitions, teachers recognized that ***complete teacher control*** over the dynamic pairing would be feasible in a **small class** if *"most students could work in individual assignments,"* (Teacher A) 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 a ***shared control between the AI system and teachers*** to maximize the support from AI systems and offload some of the orchestration tasks, as *"it would take much time to control and monitor too many students at once."* (Teacher B) 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.

VI. Discussion and Further Directions

This work delineates design alternatives to support human–AI control over dynamic transitions. Motivated by prior findings, we aimed to understand how to share control between teachers, students, and AI systems when students transition between individual and collaborative learning activities, where such transitions are not pre-planned but are instigated on the fly as the need arises, and they do not happen at the same time for all students.. Our technology probe study allowed us to evaluate shared control features and investigate teachers' and students' desires in an authentic classroom environment before moving to the full development of a co-orchestration tool. Teachers and students participated in the orchestration of dynamic transitions by experiencing a specific policy (student-, teacher- or AI system-controlled pairing policy) to transition from an individual to collaborative activity. This work raises several issues for discussion.

A. Classroom characteristics

Findings from this study suggest the need for a form of *hybrid human-AI control* which is shared between students, teachers, and AI systems and which is sensitive to classroom characteristics and social dynamics [40]–[42]. Learning performance in K-12 classrooms is driven by personal and socio-contextual factors [43]. Therefore, educational technology (EdTech) tools, such as orchestration systems, should factor in these classroom characteristics so teachers can tailor learning experiences. It is worth noting that the classroom differentiation presented in Section V illustrates how these characteristics may affect the boundaries and limits of shared control in the co-orchestration of dynamic transitions. Albeit, we cannot generalize these results nor prove statistical differences among classes due to the small sample size. For instance, teachers envisaged how the shared control with students depend on classroom size, learning performance, gender, and social dynamics. This perspective is in line with prior research exploring how orchestration might be shared between teachers, students, and technologies, to make the best use of teachers' limited time [10], [13], [38], [44]. While prior findings suggest the need for shared control over pairing decisions between teachers and AI systems (cf. [13]), our work suggests a need for all participants, students, teachers and the AI system to take a role in the co-orchestration of dynamic transitions (cf. [10], [44]) – although the ideal balance of control may vary based on classroom characteristics and dynamics, as described in the examples presented in Section V.

[4] As published by the National Center for Education Statistics [39]

B. Tensions Between Students' and Teachers' Desires

Findings from this study also surfaced some tensions between students' and teachers' desires. While students' desires pointed to greater control when selecting their partners, teachers would prefer to be able to override these selections. Also, students and teachers valued the role of the AI system when automatically matching students who were not progressing in their skill levels. Nonetheless, teachers still want to be able to override AI-system decisions or give the AI system some constraints that are shaped by the teacher's expertise and classroom dynamics. These tensions are aligned with prior findings [13] Co-orchestration systems' features should be aligned with teachers' expectations and beliefs but also should account for students' and AI participation. Future investigations should address how co-orchestration support might be designed to help teachers productively share control over orchestration tasks without leaving students and AI systems out of the co-orchestration task.

C. Adaptivity and adaptability in Co-Orchestration Systems

Moving beyond prior work on the design of co-orchestration support, which focused on understanding general needs among teachers and students, the current study surfaced variations in teacher and student needs across different classes. The current work points to the need for adaptivity and/or adaptability [45] for different classroom contexts, teacher preferences, and students' prior knowledge [8], [40], which are relevant features in AI systems.

One example of ***adaptability*** in a co-orchestration system could be enabling teachers to select the best pairing policy based on their goals, needs, and classroom dynamics (i.e., classroom characteristics). In a small size class, teachers may allow students to choose which peers they would like to work with – perhaps supported by the AI system. By contrast, in a large size class, teachers might choose to have the AI system take more control over the pairing decisions (perhaps within constraints pre-configured by the teacher - cf. [41]). Another example of adaptability could be enabling the teacher to override AI or student decisions when deemed necessary. As indicated in prior work, teachers may wish to be able to prevent or override students or AI's decisions on a case-by-case basis, but without necessarily being the point of responsibility for initiating peer tutoring activities [10], [44]. Also, as reported by Yang and colleagues, the system can adjust thresholds in order to maximize students' engagement in peer tutoring according to classroom dynamics [36]. For example, in a *high-achieving classroom*, as suggested by teachers, the system may pair up students of similar knowledge level for the tutoring activity, knowing that the pairing assignment is for social purposes. Concerning the pairing policies adopted in this study, further research should also explore the feasibility of the pairing algorithms simulated in the AI system policy. For instance, in the AI system policy, we only considered skill mastery as the primary strategy for teaming up students. However, other pairing strategies and data can be used for optimizing collaboration (e.g., [20]–[22]). For example, the AI system can learn from past interactions using historical data and predictive analytics [22].

In addition to adaptability, our findings also suggest promise for co-orchestration systems that are ***adaptive*** to teachers' and classrooms' needs, as discussed above. 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). Further work should examine design features to provide a fluid scaffolding of these dynamic transitions, so participants can quickly move from one form of instruction to another without feeling burdened. Our findings point to a need to explore the design space of context-adaptive pairing policies.

D. Hybrid Control for Human-AI Orchestration of Dynamic Transitions

Adopting a hybrid human-AI vision for supporting personalizing learning [46], we conceptualize three hybrid features towards designing systems that supports a shared control between the AI system, teachers, and students:

Teacher/student assistance. As suggested in [46], at this level, the teacher/student has full control of the orchestration tasks, and the AI system provides supportive information. Based on our findings, we suggest that the AI system could provide supporting data/information to teachers and students, such as current students' status and behavior with the system (i.e., errors made and skills in Lynnette). For example, in *struggling classrooms*, teachers (c.f., [17]) may want to examine students' errors, and skill mastery from Lynnette to make informed decisions for maximizing pairing assignments outcomes [47]. These scenarios can be applied in *small size* or *high-achieving* classrooms. As suggested by students during the workshops (see Section V, RQ1), they could be presented with data/information to select their partner, for example, based on tutoring skills.

Partial AI Automation. In this level, teachers and students monitor responses from the AI system, and the AI system controls specific tasks [46]. An illustrative example could be that the AI system (i.e., Lynnette) diagnoses students' status (e.g., off-task, gaming the system, among others [10]) and *advises* which students may need help [48]. This advice information can be consulted either by the teacher or the student. For example, as suggested by teachers, when exposed to the AI system-controlled pairing policy, they would like to have the option to reject/accept the advice from the AI system (c.f., [17]), specifically in *struggling classrooms*. On the other hand, students may also have the control to accept/reject the AI system's advice. While students or teachers did not mention this scenario, we suggest this may provide better support in *high-achieving classrooms*, where students value their learning [49].

Conditional AI automation. In this level, The AI system controls most of the orchestration tasks and advises teachers or students when control is needed (low human control). For example, the AI system could notify teachers and students that a new pairing activity has been initiated. Information may be shown to the teacher and students about the decisions of the AI system (i.e., based on some criteria, skills, behaviors [17]), and the teacher can override the AI system decisions, as suggested by teachers during the interviews. According to teachers' desires, this scenario can be realistic in *struggling* and *large size* classrooms. Also, according to teachers, it is not expected that students have a role in stopping AI system decisions. Earlier findings reported that, in practice, even when teachers are given the option to allow greater student control over some orchestration tasks, they did not share control with students [38].

E. Limitations and Further Directions

This work, however, is subject to several limitations. The major limitation is related to the limited data we could get from this authentic scenario. Due to the limited exposure to the pairing policies that teachers and students experienced, these results are to be taken as illustrative examples of the type of human-AI shared control and support that orchestration technologies should consider. Due to the low data sample of pairing information, We did not run any statistical tests to test for relevant differences in students' learning experiences or outcomes that might result from these policies. The statistical results presented in Section V (specifically in Table II and Figure 5) aimed to understand students' behavior during the exposure to pairing policies. The statistical tests were run using (individual) student data, using a similar approach as in [36].

It is worth noting that design techniques for human-AI systems, such as the ones applied in this study (i.e., storyboards and probe studies), could help explore, understand and prompt human needs before moving to full deployment. This study's storyboards and results served as a baseline for other studies (see [17], [31], [42] for details). These other studies have built upon our findings and have investigated further design concepts and tool features with more teachers and students, focusing on the teacher-AI hybrid co-orchestration. For instance, in [31], we explored design features and different levels of control for teacher-AI co-orchestration (as suggested in Section V- RQ2 and Section VI) with teachers. Moreover, in [42], we delved deeper into the design of a teacher-AI co-orchestration tool from co-designing with teachers. Finally, in [17], we presented the results of the first testing of a teacher-AI co-orchestration ecosystem with teachers and students in authentic classrooms.

VII. Conclusion

This 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 helped us gather experimentally 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 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 a hybrid human-AI control of orchestration activities, needs for further research on how particular orchestration tasks can best be balanced between teachers, students, and AI systems, and 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.

Acknowledgments

We thank the reviewers for helping improve the manuscript with their thoughtful comments. We also thank all participating teachers and students. This work was partially supported by Grant No. 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.

References

[1] P. Dillenbourg, "Framework for integrated learning," 2004. [Online]. Available: https://telearn.archives-ouvertes.fr/hal-00190107

[2] A. Kothiyal, S. Murthy, and S. Iyer, "Think-pair-share in a large cs1 class: does learning really happen?" in *Proc. of the 2014 conference on Innovation & technology in computer science education*. ACM, 2014, pp. 51–56.

[3] E. Aronson, *The Jigsaw classroom*. Sage Publications, 1978.

[4] D. Mullins, N. Rummel, and H. Spada, "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*, vol. 6, no. 3, pp. 421–443, 2011.

[5] P. Dillenbourg, "Design for classroom orchestration," *Computers & Education*, vol. 69, pp. 485–492, 2013.

[6] C. Kaendler, M. Wiedmann, N. Rummel, and H. Spada, "Teacher competencies for the implementation of collaborative learning in the classroom: A framework and research review," *Educational Psychology Review*, vol. 27, no. 3, pp. 505–536, 2015.

[7] R. Martinez-Maldonado, A. Clayphan, K. Yacef, and J. Kay, "Mtfeedback: Providing notifications to enhance teacher awareness of small group work in the classroom," *IEEE Transactions on Learning Technologies*, vol. 8, no. 2, pp. 187–200, 2015.

[8] M. Tissenbaum and J. Slotta, "Supporting classroom orchestration with real-time feedback: A role for teacher dashboards and real-time agents," *Intl. Journal of Computer-Supported Collaborative Learning*, vol. 14, no. 3, pp. 325–351, 2019.

[9] A. van Leeuwen, "Learning analytics to support teachers during synchronous cscl: Balancing between overview and overload," *Journal of Learning Analytics*, vol. 2, no. 2, pp. 138–162, 2015.

[10] K. Holstein, B. M. McLaren, and V. Aleven, "Co-designing a real-time classroom orchestration tool to support teacher–ai complementarity," *Journal of Learning Analytics*, vol. 6, no. 2, pp. 27–52, 2019.

[11] J. W. Schofield, R. Eurich-Fulcer, and C. L. Britt, "Teachers, computer tutors, and teaching: The artificially intelligent tutor as an agent for classroom change," *American Educational Research Journal*, vol. 31, no. 3, pp. 579–607, 1994.

[12] K. Holstein, B. M. McLaren, and V. Aleven, "Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms," in *Proc. of the 7th Intl. conference on learning Analytics and knowledge*. ACM, 2017, pp. 257–266.

[13] J. K. Olsen, N. Rummel, and V. Aleven, "Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom," *Intl. Journal of Artificial Intelligence in Education*, vol. 31, no. 1, pp. 24–56, 2021.

[14] M. Mavrikis, E. Geraniou, S. Gutierrez Santos, and A. Poulovassilis, "Intelligent analysis and data visualisation for teacher assistance tools: The case of exploratory learning," *British Journal of Educational Technology*, vol. 50, no. 6, pp. 2920–2942, 2019.

[15] A. van Leeuwen and N. Rummel, "Orchestration tools to support the teacher during student collaboration: a review," *Unterrichtswissenschaft*, vol. 47, no. 2, pp. 143–158, 2019.

[16] K. Holstein, G. Hong, M. Tegene, B. M. McLaren, and V. Aleven, "The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers," in *Proc. of the 8th Intl. conference on learning Analytics and knowledge*. ACM, 2018, pp. 79–88.

[17] K. Yang, Z. Lu, V. Echeverria, J. Sewall, L. Lawrence, N. Rummel, and V. Aleven, "Technology ecosystem for orchestrating dynamic transitions between individual and collaborative ai-tutored problem solving," in *Proc. of the Intl. conference on artificial intelligence in education - AIED 2022*. Springer, 2022, pp. 1–6.

[18] K. Holstein, B. M. McLaren, and V. Aleven, "Spacle: investigating learning across virtual and physical spaces using spatial replays," in *Proc. of the 7th Intl. conference on learning Analytics and knowledge*. ACM, 2017, p. 358–367.

[19] S. Amara, J. Macedo, F. Bendella, and A. Santos, "Group formation in mobile computer supported collaborative learning contexts: A systematic literature review," *Journal of Educational Technology & Society*, vol. 19, no. 2, pp. 258–273, 2016.

[20] N. Maqtary, A. Mohsen, and K. Bechkoum, "Group formation techniques in computer-supported collaborative learning: A systematic literature review," *Technology, Knowledge and Learning*, vol. 24, no. 2, pp. 169–190, 2019.

[21] I. Srba and M. Bielikova, "Dynamic group formation as an approach to collaborative learning support," *IEEE transactions on learning technologies*, vol. 8, no. 2, pp. 173–186, 2014.

[22] I. Amarasinghe, D. Hernández-Leo, and A. Jonsson, "Data-informed design parameters for adaptive collaborative scripting in across-spaces learning situations," *User Modeling and User-Adapted Interaction*, vol. 29, no. 4, pp. 869–892, 2019.

[23] M. Tissenbaum, M. Lui, and J. D. Slotta, "Co-designing collaborative smart classroom curriculum for secondary school science." *J. Univers. Comput. Sci.*, vol. 18, no. 3, pp. 327–352, 2012.

[24] V. Aleven, B. M. McLaren, J. Sewall, and K. R. Koedinger, "A new paradigm for intelligent tutoring systems: Example-tracing tutors," *Intl. Journal of Artificial Intelligence in Education*, vol. 19, no. 2, pp. 105–154, 2009.

[25] K. R. Koedinger, K. Cunningham, A. Skogsholm, B. Leber, and J. Stamper, "A data repository for the edm community: The pslc datashop," *Handbook of Educational Data Mining*, 2010.

[26] E. Walker, S. Walker, N. Rummel, and K. R. Koedinger, "Using problem-solving context to assess help quality in computer-mediated peer tutoring," in *Proc. of the Intl. conference on intelligent tutoring systems - ITS 2010*. Springer, 2010, pp. 145–155.

[27] Y. Long and V. Aleven, "Enhancing learning outcomes through self-regulated learning support with an open learner model," *User Modeling and User-Adapted Interaction*, vol. 27, no. 1, pp. 55–88, 2017.

[28] M. Waalkens, V. Aleven, and N. Taatgen, "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*, vol. 60, no. 1, pp. 159–171, 2013.

[29] E. Walker, N. Rummel, and K. R. Koedinger, "Adaptive intelligent support to improve peer tutoring in algebra," *Intl. Journal of Artificial Intelligence in Education*, vol. 24, no. 1, pp. 33–61, 2014.

[30] M. Wessner and H.-R. Pfister, "Group formation in computer-supported collaborative learning," in *Proc. of the 2001 Intl. ACM SIGGROUP conference on supporting group work*, 2001, Conference Proc., pp. 24–31.

[31] K. B. Yang, L. Lawrence, V. Echeverria, B. Guo, N. Rummel, and V. Aleven, "Surveying teachers' preferences and boundaries regarding human-ai control in dynamic pairing of students for collaborative learning," in *Proc. of the European Conference on Technology Enhanced Learning - ECTEL'21*. Springer, 2021, pp. 260–274.

[32] H. Hutchinson, H. Hansen, N. Roussel, B. Eiderbäck, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, and H. Evans, "Technology probes: inspiring design for and with families," in *Proc. of the conference on Human factors in computing systems - CHI '03*. ACM Press, 2003, p. 17.

[33] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User modeling and user-adapted interaction*, vol. 4, no. 4, pp. 253–278, 1994.

[34] J. B. Stang and I. Roll, "Interactions between teaching assistants and students boost engagement in physics labs," *Physical Review Special Topics-Physics Education Research*, vol. 10, no. 2, p. 020117, 2014.

[35] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," *Qualitative health research*, vol. 15, no. 9, pp. 1277–1288, 2005.

[36] K. B. Yang, V. Echeverria, X. Wang, L. Lawrence, K. Holstein, N. Rummel, and V. Aleven, "Exploring policies for dynamically teaming up students through log data simulation." *Intl. Educational Data Mining Society*, 2021.

[37] J. E. Beck and Y. Gong, "Wheel-spinning: Students who fail to master a skill," in *Proc. of the Intl. conference on artificial intelligence in education - AIED 2013*. Springer, 2013, pp. 431–440.

[38] M. Sharples, "Shared orchestration within and beyond the classroom," *Computers & Education*, vol. 69, pp. 504–506, 2013.

[39] S. Taie and R. Goldring, "Characteristics of public and private elementary and secondary school teachers in the united states: Results from the 2017-18 national teacher and principal survey. first look. nces 2020-142." *National Center for Education Statistics*, 2020.

[40] L. Lawrence and E. Mercier, "Co-design of an orchestration tool: Supporting engineering teaching assistants as they facilitate collaborative learning." *Timeline*, vol. 2017, 2017.

[41] J. K. Olsen, N. Rummel, and V. Aleven, "Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom," *Intl. Journal of Artificial Intelligence in Education*, vol. 31, no. 1, pp. 24–56, 2021.

[42] L. Lawrence, B. Guo, K. Yang, V. Echeverria, Z. Kang, V. Bathala, C. Li, W. Huang, V. Aleven, and N. Rummel, "Co-designing ai-based orchestration tools to support dynamic transitions: Design narratives through conjecture mapping," *Proc. of the Intl. Society of the Learning Sciences conference - ISLS'22*, 2022.

[43] J. Lee and V. J. Shute, "Personal and social-contextual factors in k–12 academic performance: An integrative perspective on student learning," *Educational psychologist*, vol. 45, no. 3, pp. 185–202, 2010.

[44] L. P. Prieto, M. H. Dlab, I. Gutiérrez, M. Abdulwahed, and W. Balid, "Orchestrating technology enhanced learning: a literature review and a conceptual framework," *Intl. Journal of Technology Enhanced Learning*, vol. 3, no. 6, p. 583, 2011.

[45] K. Holstein, V. Aleven, and N. Rummel, "A conceptual framework for human–ai hybrid adaptivity in education," in *Proc. of the Intl. conference on artificial intelligence in education - AIED 2020*. Springer, 2020, pp. 240–254.

[46] I. Molenaar, *Personalisation of learning: Towards hybrid human-AI learning technologies*. OECD Publishing, 2021, p. 57.

[47] V. Aleven, J. Blankestijn, L. Lawrence, T. Nagashima, and N. Taatgen, "A dashboard to support teachers during students' self-paced ai-supported problem-solving practice," in *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption*, I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, and T. Farrell, Eds. Cham: Springer Intl. Publishing, 2022, p. 16–30.

[48] A. Van Leeuwen and N. Rummel, "Comparing teachers' use of mirroring and advising dashboards," in *Proc. of the 10th Intl. conference on learning analytics and knowledge*. ACM, 2020, pp. 26–34.

[49] J. A. Ritchotte, D. Suhr, N. F. Alfurayh, and A. K. Graefe, "An exploration of the psychosocial characteristics of high achieving students and identified gifted students: Implications for practice," *Journal of Advanced Academics*, vol. 27, no. 1, pp. 23–38, 2016.