Received: 28 February 2023 | Accepted: 24 July 2023

DOI: 10.1111/bjet.13372

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process

LuEttaMae Lawrence[1] | Vanessa Echeverria[2,3] | Kexin Yang[4] | Vincent Aleven[4] | Nikol Rummel[5]

[1]Instructional Technology and Learning Sciences, Utah State University, Logan, Utah, USA

[2]Faculty of Information Technology, Monash University, Clayton, Victoria, Australia

[3]Facultad de Ingeniería en Electricidad y Computación, Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador

[4]Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[5]Educational Psychology, Ruhr-University Bochum, Bochum, Germany

Correspondence
LuEttaMae Lawrence, Utah State University, Logan, UT, USA.
Email: lu.lawrence@usu.edu

Funding information
National Science Foundation, Grant/Award Number: 1822861

Abstract

Artificial intelligence (AI) can enhance teachers' capabilities by sharing control over different parts of learning activities. This is especially true for complex learning activities, such as dynamic learning transitions where students move between individual and collaborative learning in un-planned ways, as the need arises. Yet, few initiatives have emerged considering how shared responsibility between teachers and AI can support learning and how teachers' voices might be included to inform design decisions. The goal of our article is twofold. First, we describe a secondary analysis of our co-design process comprising six design methods to understand how teachers conceptualise sharing control with an AI co-orchestration tool, called *Pair-Up*. We worked with 76 middle school math teachers, each taking part in one to three methods, to create a co-orchestration tool that supports dynamic combinations of individual and collaborative learning using two AI-based tutoring systems. We leveraged qualitative content analysis to examine teachers' views about sharing control with *Pair-Up*, and we describe high-level insights about the human-AI interaction, including control, trust, responsibility, efficiency, and accuracy. Secondly, we use our results as an example showcasing how human-centred learning analytics can be applied to the design of human-AI technologies and share reflections for human-AI technology designers regarding the methods that might be fruitful to elicit teacher feedback and ideas. Our findings illustrate the design of a novel co-orchestration tool to facilitate the transitions between individual and collaborative learning and highlight considerations and reflections for designers of similar systems.

KEYWORDS

co-orchestration, design methods, human-AI, human-centred design, teacher-centred

Practitioner notes

What is already known about this topic:

- Artificial Intelligence (AI) can help teachers facilitate complex classroom activities, such as having students move between individual and collaborative learning in unplanned ways.
- Designers should use human-centred design approaches to give teachers a voice in deciding what AI might do in the classroom and if or how they want to share control with it.

What this paper adds:

- Presents teacher views about how they want to share control with AI to support students moving between individual and collaborative learning.
- Describes how we adapted six design methods to design AI features.
- Illustrates a complete, iterative process to create human-AI interactions to support teachers as they facilitate students moving from individual to collaborative learning.

Implications for practice:

- We share five implications for designers that teachers highlighted as necessary when designing AI-features, including control, trust, responsibility, efficiency and accuracy.
- Our work also includes a reflection on our design process and implications for future design processes.

INTRODUCTION

Human-AI partnerships aim to enhance human capabilities and promote synergy between human actions and automation through predefined rules and structures (Dubey et al., 2020; Shneiderman, 2022). The automation includes AI techniques that analyse and predict student interactions to develop recommendations to inform instruction, specifically, specifically, humans define these processes and make decisions about the AI's level of control (Fang et al., 2023). The new research direction to intentionally combine human and AI capabilities brings novel design challenges into educational contexts. Designers must now consider the shared roles between humans and AI to effectively support teaching and learning practices. Several researchers have conceptualised how to foster collaboration between humans and AI. For instance, Molenaar's framework (2022) aimed to characterise human-AI interactions according to six levels of automation, from zero to full automation. Similarly, Holstein et al. (2020) proposed a hybrid adaptivity framework to define shared adaptivity with human-AI regarding perceptual, actions and decision capabilities. These initiatives highlight the need for a deeper understanding of the interaction between humans and AI in educational contexts. Designing human-AI technologies contains the difficult challenges of fostering productive interactions and enabling trust, transparency, explainability and responsibility (Ozmen Garibay et al., 2023).

By adapting methods from human-computer interaction (HCI) to meaningfully include stakeholders in the design technology (Buckingham Shum et al., 2019), the field of human-centred learning analytics (HCLA) has addressed similar challenges. However, researchers argued the same methods do not transfer neatly to AI because of additional complexities: understanding AI, explaining AI design ideas, and communicating AI unpredictability (Xu et al., 2023; Yang et al., 2020). Xu et al. (2023) explained traditional HCI methods need to be adapted to address some of these challenges. Simultaneously, they call for HCLA to continue adapting methods to maximise the inclusion of stakeholders and foster better interactions with AI (Ozmen Garibay et al., 2023; Xu et al., 2023). There are few examples of how researchers have codesigned AI-based tools with teachers (eg, Holstein et al., 2019). We build on the work of Holstein et al. (2019) to explore how teachers can codesign AI-based orchestration tools that distribute control during unplanned learning activities like dynamic pairing, where students move between individual and collaborative learning, described in detail below.

Our research had two purposes. First, we describe a secondary analysis, which serves the purpose of answering our research question: *How do teachers conceptualise sharing control with an AI co-orchestration tool during dynamic pairing?* Second, we use our results to illustrate how HCLA can be applied to the design of human-AI technologies. Subsequently, our article is outlined as follows: first, we review the literature on co-orchestration and dynamic learning transitions; next, we describe our methods, including the AI context, the six design methods we used, and our analytic approach; after, we share the results from our research question; lastly, in the discussion, we summarise our results related to our research question, highlight implications for this work and reflect on our HCLA process.

CO-ORCHESTRATING DYNAMIC LEARNING TRANSITIONS

Teachers manage many simultaneous factors in classroom environments to facilitate meaningful student learning. Also known as classroom orchestration, this approach refers to the moves teachers undertake to manage tools and students in real time (Dillenbourg & Jermann, 2010). Orchestrating classroom learning becomes vastly more taxing when learning approaches are complex. Dynamic transitions, for instance, require teachers to manage students working on individual and collaborative learning and identify when they should switch between learning activities (Echeverria et al., 2023). Leveraging the knowledge learning instruction (KLI) framework (Koedinger et al., 2012), collaborative learning offers opportunities for mutual elaboration and coconstruction of knowledge. In contrast, individual learning may promote induction and refinement as learning mechanisms (cf. Mullins et al., 2011). For example, Olsen et al. (2015) showed that combining individual and collaborative learning can be more effective for students than either activity alone. Specifically, we define dynamic transitions as having three components: (1) monitoring student learning skills, status (eg, doing well) and progress; (2) having students transition when the need arises and (3) pairing students on the fly (Yang et al., 2023). In educational practice, combining individual and collaborative learning activities is common (eg, think-pair-share), as each learning activity has unique benefits. Such pairing processes, like think-pair-share, are ones that teachers know how to handle. What makes dynamic pairing processes different is that students are working on different tasks on different timelines and are being paired with different students on the fly, meaning students switch back-and-forth between individual and collaborative learning at unplanned and somewhat unpredictable times (Olsen et al., 2015). The dynamic element introduces a new opportunity to adapt students' learning in real time to account for their unique needs. While we hypothesise that dynamic learning transitions can be effective for students' learning it is empirically unproven. We can a priori see advantages but do not yet have scientific studies that have rigorously tested them. One challenge in implementing and studying this form of learning is that dynamic learning transitions are a complex pedagogical approach for teachers to orchestrate.

To benefit individual student learning, teachers require support identifying moments to transition between individual and collaborative learning. To scaffold teacher implementation of dynamic learning transitions, AI can support teachers as they orchestrate learning transitions, also known as co-orchestration. Co-orchestration involves shared decision making and coregulated behaviour and interaction in the classroom (Holstein & Olsen, 2023). For co-orchestration to be effective, the collaborative features between teachers and AI need to be grounded in the needs of teachers and the realities of classrooms. Therefore, HCLA can be leveraged to position teachers as active collaborators in the design process (Dimitriadis et al., 2021; Goodyear & Dimitriadis, 2013). HCLA is a participatory design approach that has become increasingly common in Learning Analytics to account for stakeholder needs (Buckingham Shum et al., 2019). Yet, when applying HCLA to AI technologies, new challenges emerge such as understanding what AI can do, uncertainty regarding the role of AI in the classroom, and difficulty designing interactions to foster collaborative decision-making processes (Yang et al., 2020). Thus, methods to design AI require specific adaptations to engage stakeholders in meaningful ways (Ozmen Garibay et al., 2023; Yang et al., 2020). We address these challenges by adapting six methods, described below, to design our AI co-orchestration tool with teachers.

METHODS

Learning context

The goal of our project was to create a co-orchestration tool for teachers to make decisions with AI about dynamic learning transitions in the classroom. The co-orchestration tool, *Pair-Up*, is situated in a sociotechnical classroom ecosystem, including two AI-based intelligent Tutoring Systems (ITS): *Lynette* and Adaptive Peer Tutoring Assistant (*APTA*). *Lynette* (Figure 1) is a standard ITS that supports individual learning of basic equation solving. *Lynnette* provides step-by-step guidance through adaptive hints, correctness feedback and error-specific messages, improving students' equation-solving skills in several classroom studies. *APTA* (Figure 2) extends *Lynnette's* functionality to support collaborative learning, specifically reciprocal tutoring. *APTA* is a reimplementation of an earlier system with the same name by Walker et al. (2014). When using *APTA*, two students take the role of Solver and Tutor, respectively. The Solver attempts to solve the math problem and can seek help from their partner, the Tutor. The Tutor helps the Solver through step-by-step evaluation and feedback via the chat window. *APTA* supports the student in the Tutor role with math and tutoring feedback. While we describe the larger classroom ecosystem, the focus of this paper is on how teachers wanted to share control with the co-orchestration tool, *Pair-Up*.

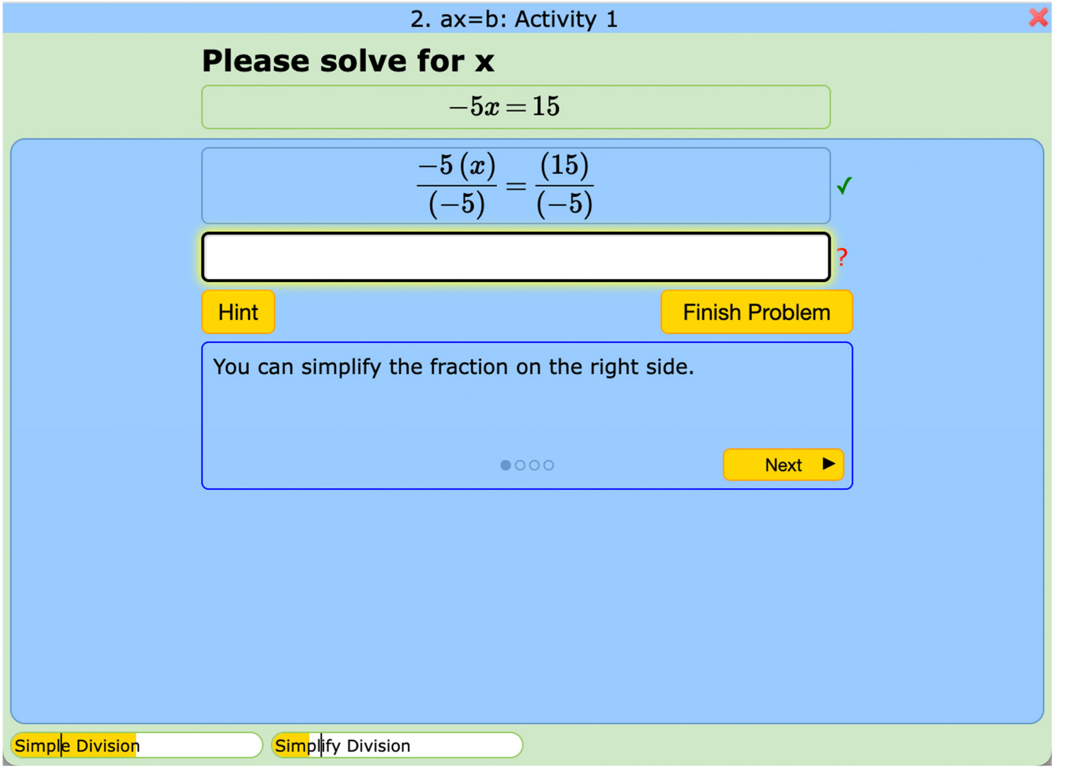

FIGURE 1 The Lynnette interface for students to solve basic equations by typing in answers, asking for hints from the AI, and seeing their progress on the skills they are working on.

The co-orchestration tool, *Pair-Up*, (Figure 3) has two key features: real time analytics of student learning status and AI-suggested pairings. *Pair-Up* (1) describes students' individual and collaborative progress and process, (2) predicts who might need collaborative support, (3) predicts who might be a beneficial partner for them and (4) advises the teacher on what pairs should work on.

First, the real time analytics use Bayesian knowledge tracing (BKT) techniques (Corbett & Anderson, 1995) from raw student-tutor interaction data to identify how students are doing (ie, doing well, struggling, system abuse, idle). *Pair-Up* also highlights what problem set students are working on and their progress within the problem set. Second, *Pair-Up* helps teachers decide which students might benefit from transitioning from one learning activity to the other. *Pair-Up t*akes advantage of tracking students' knowledge with BKT to identify students to pair. The tool leverages two algorithms: a homogeneous pairing algorithm that suggests tutors and solvers with similar mastery of the targeted math skills and a heterogeneous pairing algorithm that suggests tutors who have mastered the skill(s) necessary to support the solver with skill development. We developed the pairing algorithms through an iterative process to refine their accuracy and feasibility (Yang, Echeverria, et al., 2021). In *Pair-Up*, teachers can choose which pairing algorithm to use and whether to follow or override suggestions. Third, once the teacher selects a candidate solver, either suggested by the tool or based on their own judgement, *Pair-Up* suggests three candidates for the tutor role. Teachers can then see how each compares to the solver in terms of the skills the candidate solver is struggling with and decide which tutor to select. Lastly, teachers then select an assignment they see fit for the pair.

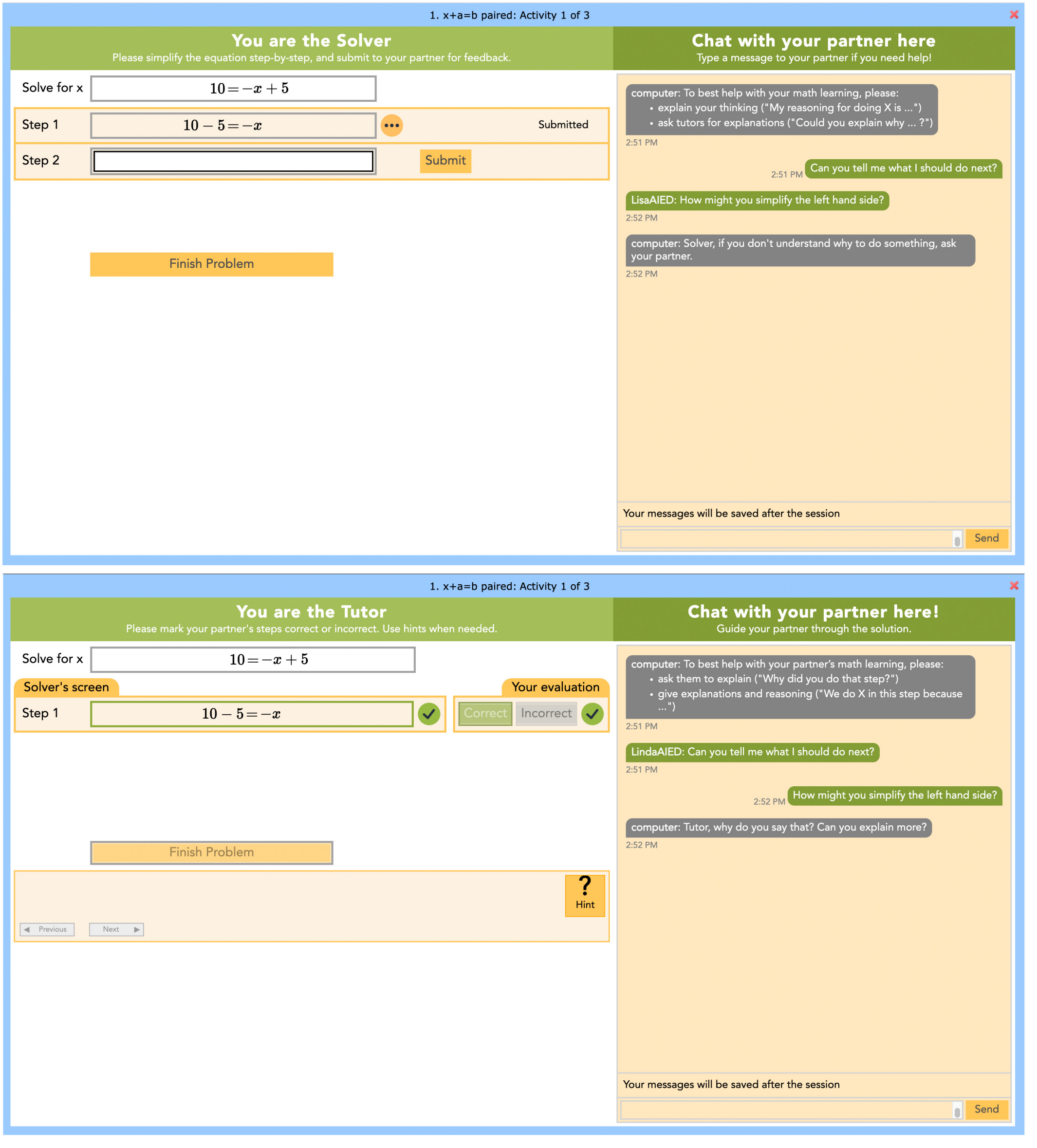

FIGURE 2 APTA interfaces. Top: Solver interface, where the student solves the equation on the left and can chat with the tutor on the right and see content hints from the AI. Bottom: Tutor interface with the solvers work on the left where they mark things as correct or incorrect and the chat on the right to provide hints to the solver and see collaboration hints from the AI.

Study design

To answer our research question, we conducted secondary analysis of our design-based research project, including six methods over 2.5 years (Table 1). Previous analysis focused on user experience design insights, features of the *Pair-Up*, and knowledge about dynamic transitions. This way, we expand on our past analyses to reflect specifically on how teachers conceptualise control across our entire design process to gain insights about how we might design for shared control in the future. The strength of collapsing across methods allows us to examine and highlight the full human centred design process. Across all six methods, we

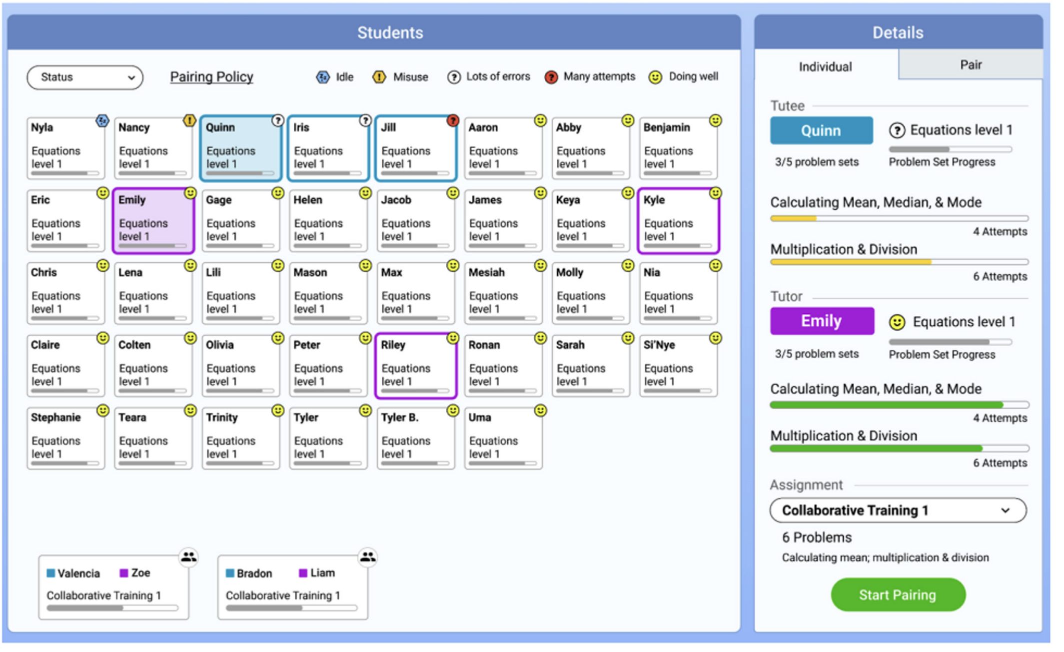

FIGURE 3 *Pair-Up* visualises all students as cards. Cards highlighted in blue represent a student who may benefit from collaboration. When selected, three students are highlighted in purple who may be a good pair. Teachers select students, and *Pair-Up* shows data about the pair in the right panel.

TABLE 1 Data collection by design method.

Design methods	Methods description	Year	Teachers	Data analysed
Classroom probe (Echeverria et al., 2023)	Wizard-of-Oz study to explore simulated interactions in authentic classrooms	Spring 2020	4	4, 30–60-minutes video interviews
Storyboard survey (Yang, Lawrence, et al., 2021)	Needs validation emphasizing the complementarity of teacher-*Pair-Up* interactions	Fall 2020	54	Qualitative survey responses
Low fidelity prototyping (Lawrence et al., 2022)	Traditional prototyping method with added imaginative prompts about intelligent functions of *Pair-Up*	Spring 2021	7	7, 1-hour videos
Mid-fidelity prototyping (Lawrence et al., 2022)	Prototyping with focus on teacher *Pair-Up* collaboration and explainability	Summer 2021	8	8, 1-hour videos
User testing (Lawrence et al., 2022)	Wizard-of-Oz study and think-aloud to simulate teacher pairing with *Pair-Up*	Fall 2021	3	3, 1-hour videos
Classroom study (Yang et al., 2023)	Testing in authentic classroom to understand teachers' real-time use of *Pair-Up*	Fall 2022	5	5, 1-hour videos

worked with a total of 76 teachers. Each teacher took part in one to three methods described below; no teacher took part in all aspects of the design process. Insights collected from each method informed succeeding designs.

Design methods

Classroom probe

To understand the design challenge and test intelligent functions, we conducted an in-person Wizard-of-Oz probe study (Hutchinson et al., 2003). During their regular classes, three, seventh-grade math teachers used the two AI-based tutoring systems, *Lynnette* and *APTA*. In total, six classes participated in the study. Each class was randomly assigned to one of three pairing conditions to transition from individual to collaborative activities, student-led, teacher-led or AI-Wizard-of-OZ-led pairing. Afterward, teachers participated in 30-minute semi-structured interviews. We prepared a set of storyboards (Figure 4) representing the three conditions to give teachers an impression of conditions they did not experience. We asked teachers to reflect on different features of the co-orchestration tool in terms of sharing

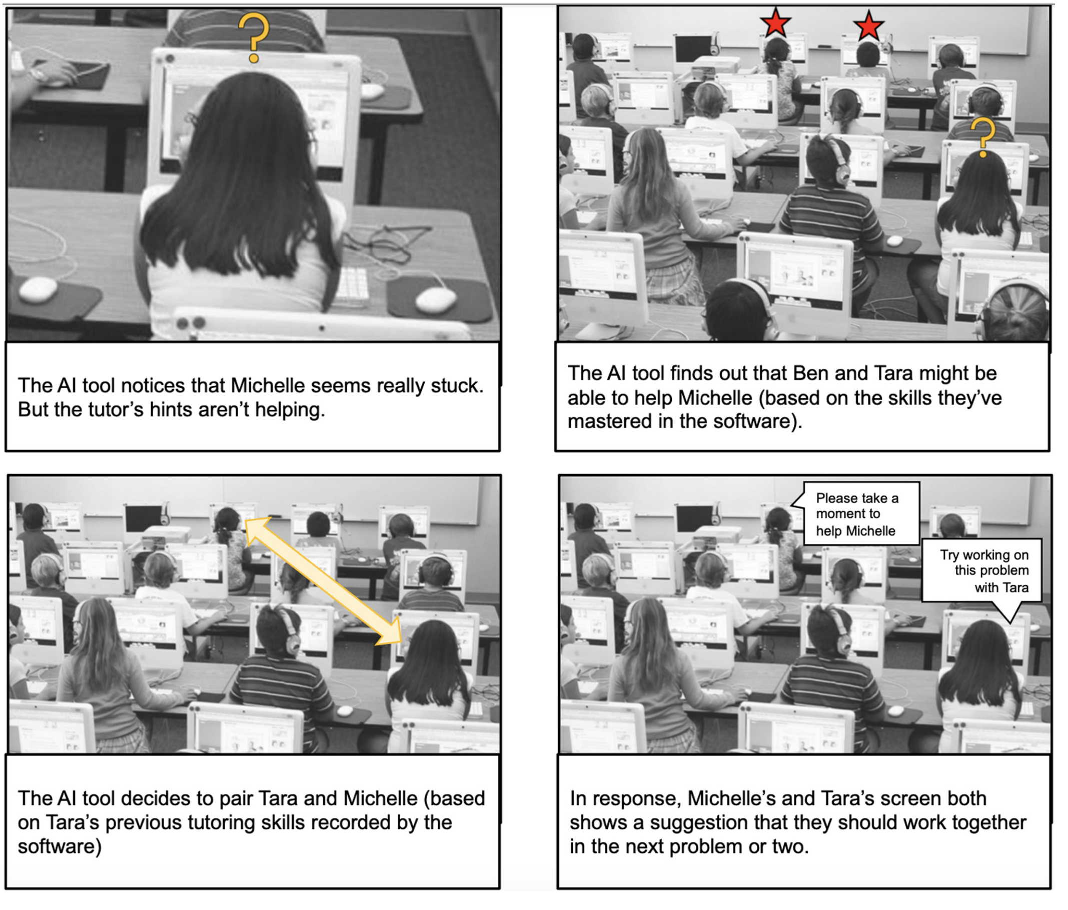

FIGURE 4 Sample of the storyboard used in the classroom probe study.

control and decision making. Our findings showed that teachers wanted hybrid control with students and the AI but also needed to account for contextual details in their classroom.

Storyboard survey

Next, we conducted a needs validation survey with storyboards (Haesen et al., 2010). Teachers reflected on different scenarios related to the co-orchestration tool. We presented five storyboards (Figure 5) illustrating concepts from teachers and outlier scenarios that pushed social boundaries about who has control over pairing (eg, *Pair-Up* always auto-pairs without the teacher). The scenarios were presented in an online survey with Likert and open-ended questions to probe teacher perceptions of the scenarios. The survey was available for seven days. Our findings revealed that teachers wanted *Pair-Up* to propose pairs and were hoping it would serve as an extra pair of eyes to evaluate student progress and notify them of what is going on instead of making decisions.

Low fidelity prototypes

Next, we co-constructed low fidelity prototypes (Buchenau & Suri, 2000). We asked teachers to comment on and generate prototypes based on how they might pair their students. Prototypes were developed in Google Slides (slides.google.com) and included the teacher's students' first names to add realism to the pairing process (Figure 6). We discussed how they might pair students and what they wanted to know to inform their pairings. We then asked teachers to imagine they had an intelligent teaching assistant to support pairing and explain how it might support them while pairing. Design ideas and prototypes that emerged from teachers were presented to subsequent teachers for iterative feedback. The design features that emerged included mechanisms for *Pair-Up* to supply suggestions, sidebars and pop-up panels that provided teachers with data about students.

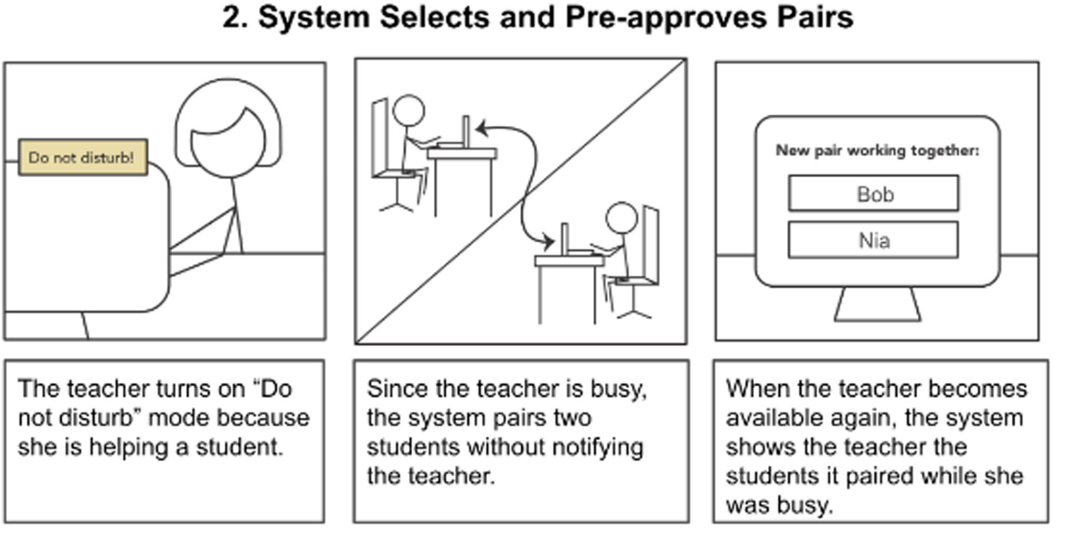

FIGURE 5 Sample of the storyboard survey.

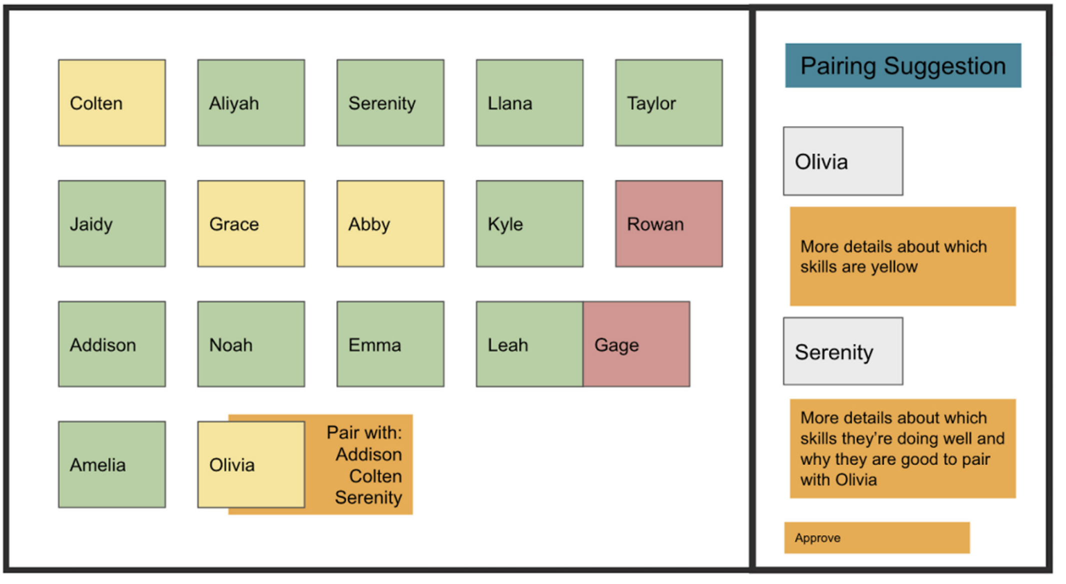

FIGURE 6 Low fidelity prototype example.

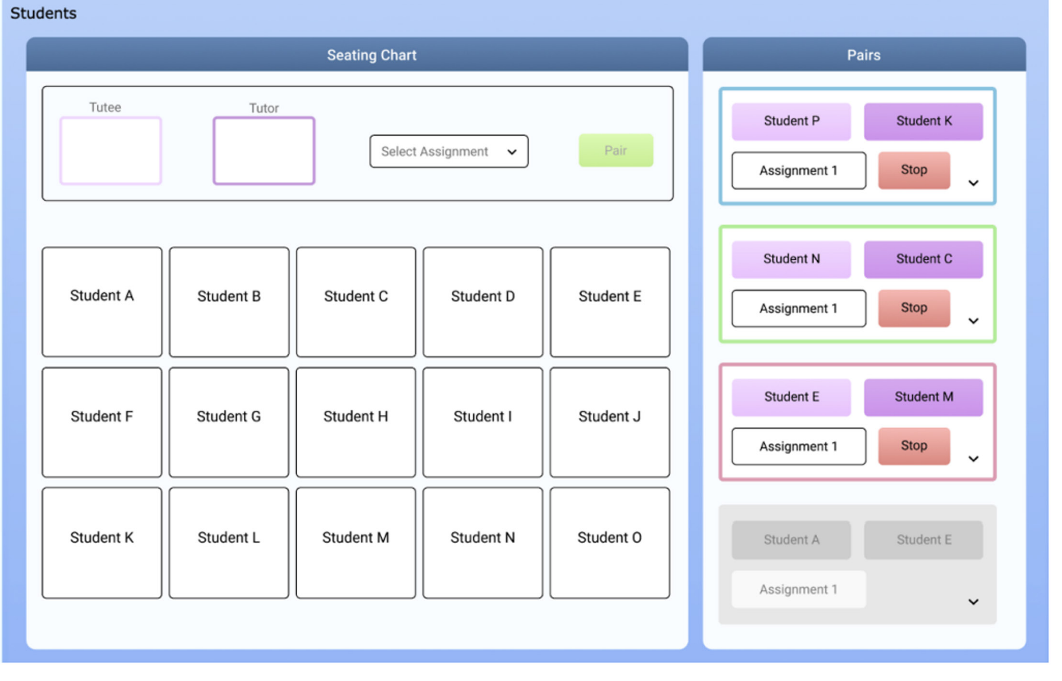

FIGURE 7 Mid-fidelity prototype example.

Mid fidelity prototype feedback

Mid-fidelity prototypes were iteratively built in Figma (figma.com) (Buchenau & Suri, 2000; Figure 7). Prototypes included student first names to provide realism. We aimed to refine the interactions while focussing on teacher-AI collaboration and explainability. We presented each prototype and probed into the following: (a) possible teacher use of the tool, potential explanations of *Pair-Up*'s recommendations and additional *Pair-Up* support for teachers. We used emergent teacher design ideas as probing questions, with minimal changes to the prototypes between teacher meetings. Our findings highlighted a need for quick, scannable information that was not overwhelming to teachers, with options to get more information if they wanted.

User testing

Next, we created a fully functional prototype for interactive user testing to simulate functionality (Figure 8). Through a video conferencing system, we gave teachers remote access to interact with *Pair-Up*. We used a Wizard-of-Oz method where the research staff worked on the tutoring software to simulate *Pair-Up*'s functionality. We prompted teachers to think aloud while pairing their students (Nielsen et al., 2002). The focus was to simulate *Pair-Up* suggestions and refine the design and interaction. After using *Pair-Up*, we presented teachers with three pairing algorithm options, explanations of each and designs for how they might be integrated. Teachers found the interface clear and informative but requested design features to customise the algorithm and presented data.

Classroom prototyping study

Lastly, we implemented *Pair-Up* to test feasibility, get initial impressions of classroom dynamics, understand student and teacher preferences, and invite their feedback (Figure 3).

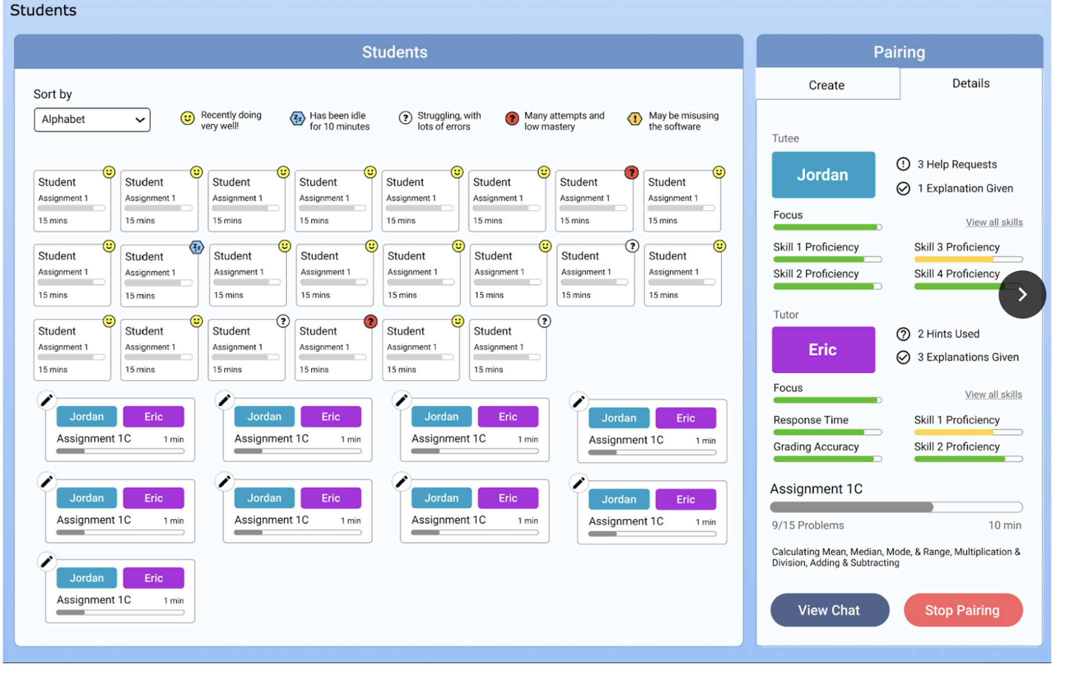

FIGURE 8 High fidelity prototype of *Pair-Up*.

We conducted an in-person classroom study in a suburban public school with five middle school math teachers and 199 students from 11 classes, one of which was a special education class. Each class participated in two lessons, each lasting 33 to 37 minutes. After the live experience, we interviewed each teacher about their perceptions of using the tool in the classroom.

Analysis

We conducted a secondary analysis to understand how teachers conceptualised sharing control with *Pair-Up*. The transcripts, videos and survey data from all six design methods were compiled for secondary analysis in MAXQDA (maxqda.com), a qualitative analysis software. Two researchers reviewed the data, identifying episodes where teachers indicated who should control the pairing process (ie, teachers, students, *Pair-Up*). In our coding, we were interested in who does the pairing; therefore, we did not include additional features related to *Pair-Up* (eg, what data are shown to teachers). We coded episodes as the unit of analysis to capture moments where teachers reflected on sharing control to capture the reasoning of how and when teachers wanted to share control and with whom. The video data included at least one turn of talk. In the survey data, episodes included individual open-ended responses. Episodes were coded into four high-level categories about who participates in the pairing process (see Table 2; Cohen's Kappa of 0.83); we discussed disagreements until reaching a consensus. We identified 457 episodes in the data set and conducted qualitative content analysis to inductively categorise episodes into themes (Mayring, 2014). Two researchers identified 45 themes, which were refined down to 14 final themes. Below, we first describe the patterns of episodes across design methods, followed by the episodes within each theme.

RESULTS

We examined teacher contributions across design methods to understand how teachers conceptualised sharing control with *Pair-Up* during dynamic transitions. Figure 9 highlights that most teachers across design methods focused on how they can share control with *Pair-Up*. Beyond that strong focus, in the initial design methods, a secondary focus was more on *All Parties* than just teachers, and we see a small shift toward more *Teacher-only* scenarios over time. Teachers in the classroom probe and storyboard survey focused on

TABLE 2 Coding scheme applied to the data.

Categories	Definition	Example
Teacher	The teacher alone controls the pairing process	The teacher makes pairings with no input from students or *Pair-Up*
Pair-Up	*Pair-Up* alone controls the pairing process	*Pair-Up*, automatically makes pairings with no input from the students or teacher
Teacher and *Pair-Up*	The teacher and *Pair-Up* collaboratively control the pairing process	*Pair-Up* recommends pairings to teachers from which the teacher can choose
All parties	Teacher, *Pair-Up* and students collaboratively control the pairing process	Students input preferences through *Pair-Up*, which are shared with the teacher

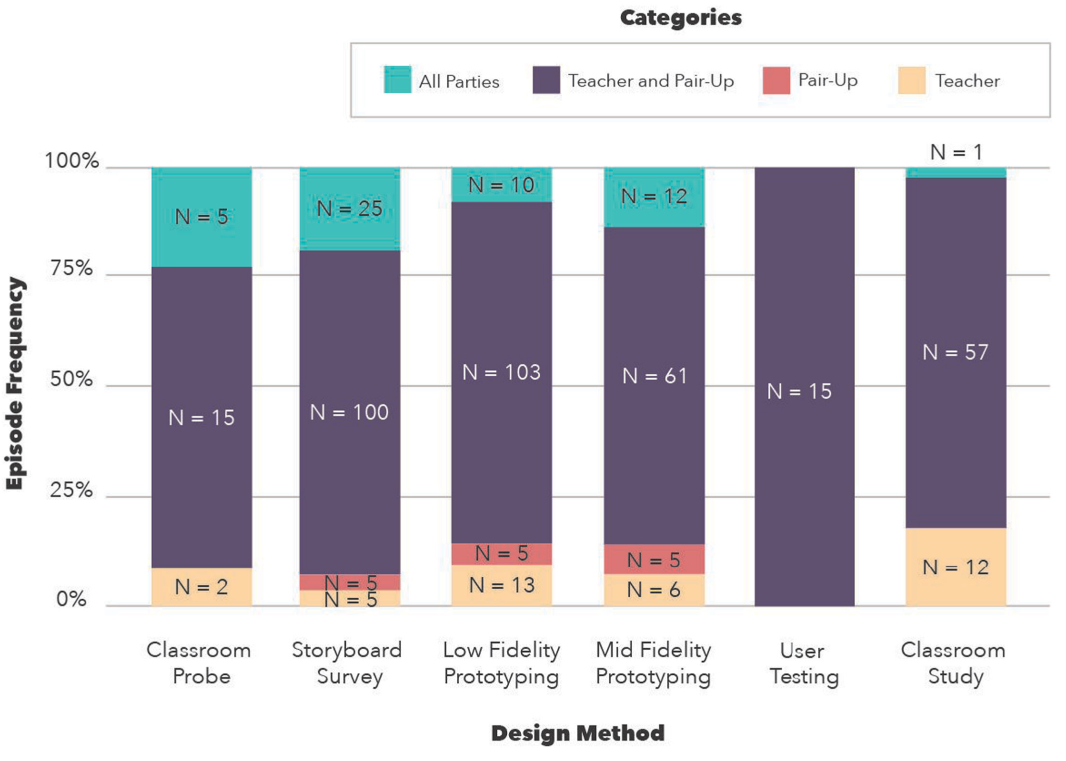

FIGURE 9 Episode frequency by category and design method.

Teacher and Pair-Up and how *All Parties* worked together. As the process progressed, the contributions pivoted away from *All Parties*. In the user testing, teachers focused only on the *Teacher and Pair-Up*, without student control. This was expected as the goal was to interact with the simulated version of *Pair-Up* and provide feedback about how teachers would work with *Pair-Up*. The classroom study followed, where teachers again focused on the *Teacher and Pair-Up* collaboration but also shifted to the *Teacher* alone. Teachers rarely wanted *Pair-Up* to completely control the pairing, with a small spike in auto-pairing only under certain circumstances, such as pairing randomly. Next, we break down episodes within each theme.

Teacher

Two themes emerged regarding teacher control (Figure 10); teachers thought they should be the one who pairs, instead of *Pair-Up* or students. Figure 10 highlights that most of these episodes emerged in classroom study. In contrast, those related to students came from low fidelity prototyping, where teachers were prompted about student roles in the pairing process.

Concerning episodes when *teachers wanted to do the pairing instead of Pair-Up* ($n=22$), some teachers were hesitant about the idea of *Pair-Up* having the control to make pairings. One teacher in the mid fidelity prototype described its limitation in not knowing what is happening in the classroom: “I think having the option to manually pair is always good. Because sometimes, there's things that don't show up in [Pair-Up], where [teachers] know who they should and shouldn't be working with.” Most of these episodes came from classroom study ($n=11$). Teachers described ignoring *Pair-Up's* recommendations because they know their

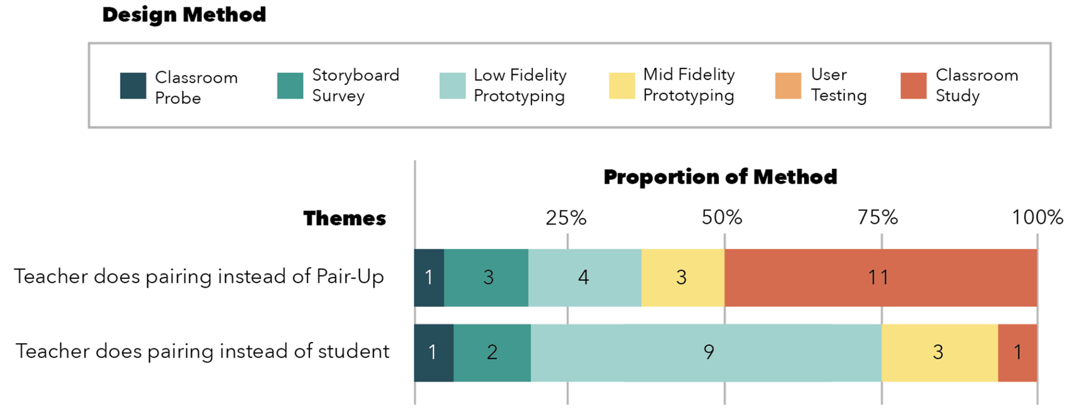

FIGURE 10 Proportion of episodes by design method in the teacher category.

students better or wanted to test out pairs who do not typically work together. In one example, a teacher shared, "I didn't really use any of the pairing suggestions for that class. Just because I've been working with them for quite a while, so like, I kind of already know the students."

Regarding episodes where the *teacher did the pairing instead of students* ($n=16$), most episodes came from low fidelity prototyping, where we had explicit prompts to discuss if students should have a say in the pairing process. Teachers explained that students would pair with their friends, deny pairings if given the option and overwhelm teachers with pairing requests. One teacher in the storyboard survey shared, "The teacher should be able to pair intentionally, but students should not be able to request in [Pair-Up]. This can lead to many problems in a middle school classroom. Misuse, hurt feelings, etc."

Pair-Up

Teachers rarely wanted *Pair-Up* to control the pairing process alone (Figure 11). Many teachers were against auto pairing without teacher awareness or control (see *Teacher and Pair-Up* below). Four teachers in the storyboard survey, low fidelity prototyping and mid-fidelity prototype found auto pairing to be effective but only under certain circumstances ($n=15$). For instance, if *Pair-Up* was randomly assigning pairs, pairing based on the number of times two students have worked together, or to support self-regulated learning. One teacher described if *Pair-Up* could learn from his pairing strategies, then over time, it could pair alone:

> If the intelligence in the system can be able to, to have the history of how I've been [pairing], so that in future, it can do it automatically, I don't have to be there. I think that will be helpful because maybe it's going to assist me. Then, it must know how I've been able to do it in the past.

Teacher and *Pair-Up*

Six themes emerged from teachers regarding the interaction with and control of *Pair-Up*. Figure 12 shows that most themes emerged across all methods except themes related to auto pairing and overriding pairing features. Most teachers strongly desired to be able to

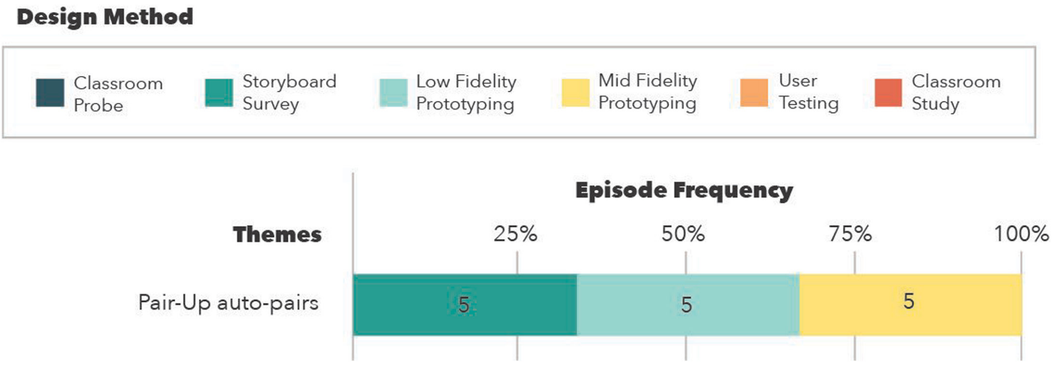

FIGURE 11 Proportion of episodes by design method in the *Pair-Up* category.

preconfigure Pair-Up before using it in the classroom ($n=116$). Teachers advocated for customisation and flexibility in *Pair-Up* to account for changing factors, including class dynamics, size, content, social dispositions, and skills. Flexibility was preferred, so teachers could set restrictive parameters based on individual needs. Throughout the initial design methods, teachers recognised *Pair-Up could share data with teachers* for two purposes: (a) to *monitor* individual and collaboration processes ($n=37$) and (b) to *make informed decisions about pairing* ($n=48$). Teachers acknowledged the need for accurate information about student knowledge (ie, skill levels, strengths, misconceptions, errors) and historical performance (eg, conceptual and tutoring skills). One teacher during the mid-fidelity prototype suggested getting information about students' idle status, off-task chat prompts and misuse of the software to intervene:

> It would be helpful if a group wasn't going well, for whatever reason, like the student isn't answering anymore, that would be a good chance to say like, 'This person needs a new tutor.' If they stop working, like if they're either idle for 10 minutes, or probably chit-chatting.

Other teachers wanted to see student strengths and what they are doing well to make effective pairings. Teachers also wanted social interactions of student conversations to act promptly. One teacher in the classroom probe expressed, "If something happens with their [students] sending an inappropriate comment or their chatting, or even if they're explaining it [the procedure] wrong, I want to be able to fix that quickly." When teachers experienced the *Pair-Up* in the classroom study, they found student skill mastery, errors, and status, which were communicated in *Pair-Up*, useful for making informed decisions about pairing.

During the initial stages of the design, teachers recognised that *Pair-Up could suggest pairings to teachers* ($n=88$). Teachers appreciate the convenience and efficiency of AI-generated pairing, as it could reduce the burden by saving time and supporting student learning. Because there may be other factors to consider when pairing students, teachers wanted to retain the final decision because they expressed reservations about relying solely on *Pair-Up* to determine the best pairings. For example, one teacher in the mid fidelity prototype shared,

> Being able to select the student and then give a few names is great, but also being able to reject that because I work with middle school students. So frequently, I'm told, you know, this girl can't work with these girls right now … So, if those three girls popped up, I can reject those [pairings].

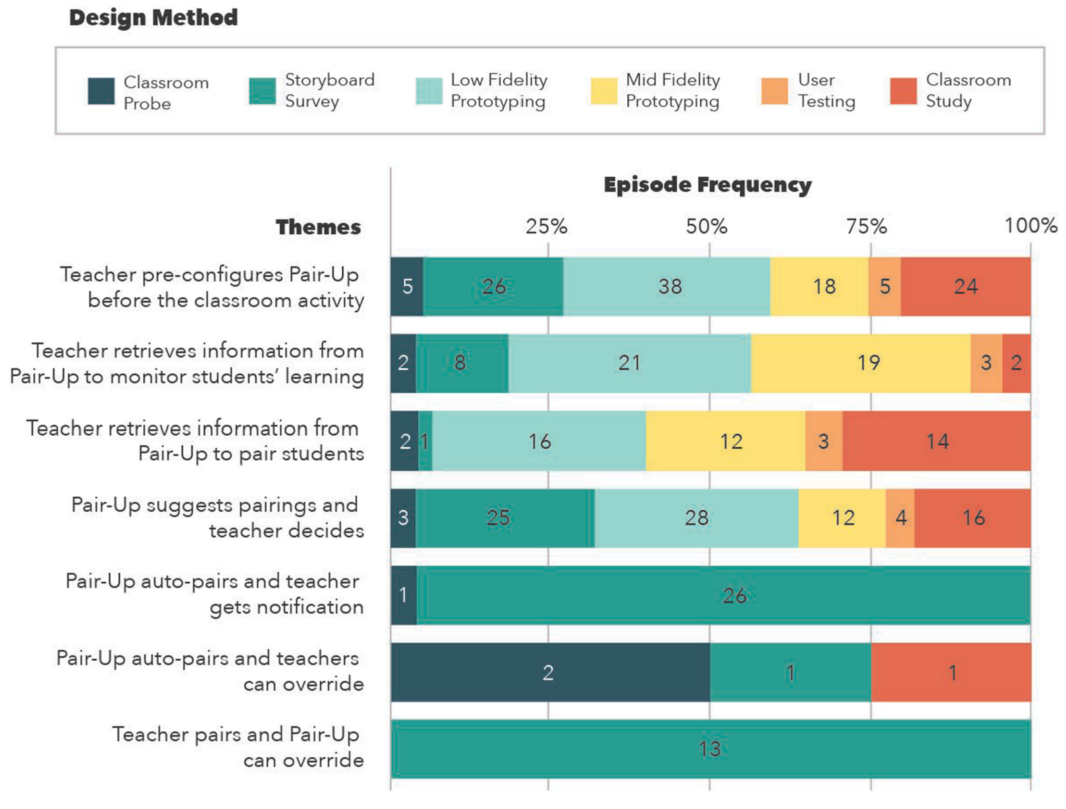

FIGURE 12 Proportion of episodes by design method in the teacher and *Pair-Up* category.

Teachers expressed the need for *Pair-Up* to provide evidence-based reasoning for its suggestions. One teacher described that she wanted *Pair-Up* to explain why it thought students were a good match:

> It might be helpful to think about why the tutors who were suggested in the first place, and then if there were a reason that made this student not a good candidate. It would be helpful to know what that is. Because even if I want to accept it, then I at least know precisely what I'm accepting.

Teachers in the classroom enjoyed having control over pairings. Four teachers wanted *Pair-Up to auto pair students*. They proposed getting *notifications* ($n=27$) to keep them informed and a feature to *override Pair-Up decisions* ($n=4$). Teachers agreed they should be informed of changes made by *Pair-Up* and desired high involvement in the pairing process. Lastly, two teachers in the storyboard survey liked the option for the *teacher to pair, but Pair-Up to override it* ($n=13$), specifically when *Pair-Up* could predict a potential problem.

All parties

Four themes emerged regarding how teachers wanted all three parties to inform pairing (Figure 13). Most themes in this category were specific to design methods, meaning these

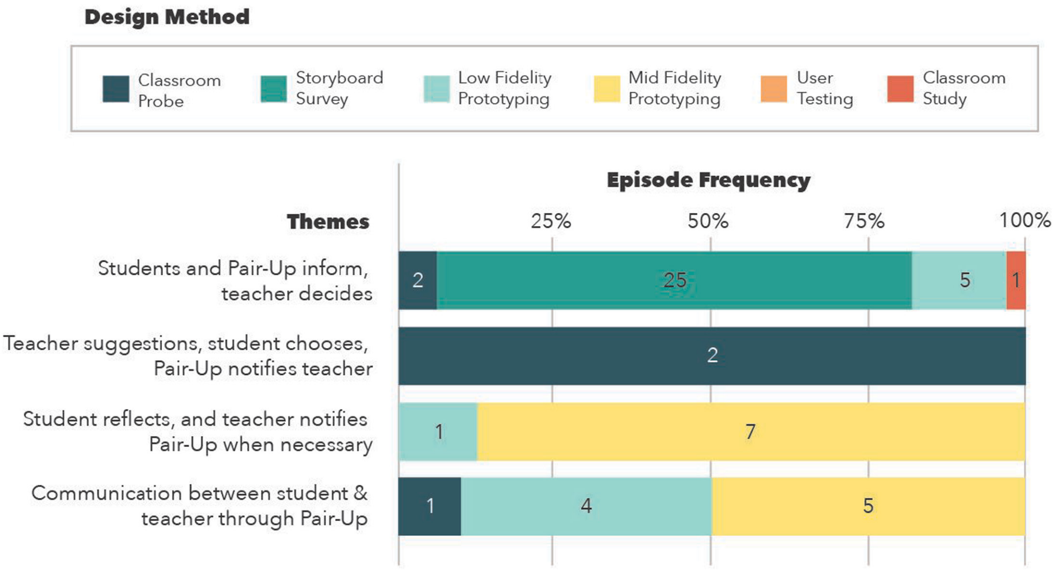

FIGURE 13 Proportion of episodes by design method in the all parties category.

themes emerged from prompts or ideas in specific design activities. Three of the themes had to do with the order in which all parties could inform the pairing. Most frequently and in the storyboard survey, teachers wanted the *students and Pair-Up to be able to inform pairing but the teacher to make the final decision* ($N=33$). Many teachers explained that students needed to have a good reason to be paired and that they would validate it with their knowledge of the pair and the data in *Pair-Up*. One teacher explained that the order in which the three parties could work together depended on the students:

> The teacher should override any other decisions … that would be like kind of the levels, like the teacher [pairs, then Pair-Up], and then student, or the teacher, and the student and [Pair-Up]…. There would be certain students where I would not want them having that control. So, it would be the teacher, [Pair-Up], then student."

She explained that some students who struggle or misbehave should not have control.

One teacher suggested *teachers could suggest pairings to students, and Pair-Up would notify whom they chose* ($n=2$). In the prototyping stages of the process, a few teachers proposed *that students could reflect on how the pairing went, notify the teacher, and then the teacher would restrict whom Pair-Up can pair in the future* ($n=8$). One teacher in the interactive prototyping shared:

> But if I had input, if I had [students' self-reflections], like from these other things like, "Today, I felt uncomfortable with that," if I'm aware of that ahead of time, I can deal with that. Give them a more polite way to say, "No, I don't want to work with them."

Lastly, several teachers expressed *using Pair-Up as a communication channel for pairing between students and the teacher* ($n=10$). Teachers emphasised that they wanted students to be able to draw the teachers' attention, for instance, through a direct message or a flag in *Pair-Up*.

DISCUSSION

Guided by our research question, we explored teacher conceptualisation and preferences of sharing control with *Pair-Up* to orchestrate dynamic transitions between individual and collaborative learning. Across design sessions, teachers preferred to collaborate with *Pair-Up* to oversee the pairing process, which reflects design factors around teacher control, while orchestrating dynamic classrooms (Dillenbourg & Jermann, 2010). Many teachers preferred to have the final say over *Pair-Up*'s recommendations with flexibility, customisation and accuracy of information about students to inform their pairing decisions. While teachers were initially keen on sharing control across the teacher, student, and *Pair-Up*, that goal shifted slightly over time. Our findings revealed that teachers rarely wanted *Pair-Up* to control the pairing process alone. Rather, teachers wanted *Pair-Up* to guide teacher attention and inform their decisions, instead of solely relying on *Pair-Up* to make independent decisions (Holstein & Olsen, 2023). This study also highlighted the lack of student control from the teacher perspective. Researchers have highlighted the importance of including students in co-orchestrating learning transitions, yet there are many open and unexplored questions regarding what that might look like given the added complexity of including student voices (Holstein & Olsen, 2023). Although we did not design for student control in our current version of *Pair-Up*, we see a need for future design methods focused on how students can have agency or control over the pairing process, in a manner acceptable to teachers. Based on these findings, we share implications for Human-AI design of similar tools.

Implications for human-AI design

High teacher control

Across methods, teachers wanted control over dynamic transitions and *Pair-Up* to make changes only with their approval. However, teachers found many features of human-AI collaboration useful, such as pairing suggestions informed by the algorithm and data to explain and inform decisions (Goodyear & Dimitriadis, 2013; Shneiderman, 2022). Designers need to consider how teachers can orchestrate classroom activities with AI to inform pedagogical decisions, while retaining teacher control.

Trust

Teachers preferred to trust their judgement and decision-making ability over the AI. In the initial design methods, teachers trusted the pairing suggestions from *Pair-Up*. However, when using *Pair-Up*, teachers felt they better understood their student's needs and relationships that might impact student collaboration. Thus, teachers may override *Pair-Up's* recommendations, meaning teachers were still hesitant to completely trust and follow AI suggestions. This teacher-AI synergy can be improved by developing a sense of trust among teachers (Dubey et al., 2020). We recognise that some scepticism toward AI is good, as teachers can make informed classroom decisions independent of AI. Nonetheless, once developing the self-efficacy to work with AI, teacher decision making may be enhanced, streamlined or supported.

Responsibility

Even when collaborating with *Pair-Up*, teachers felt responsible for classroom orchestration decisions. This mirrors findings in the literature (Molenaar, 2022), where humans and

AI share control, but humans are responsible for any actions in the classroom. Designers of AI technologies must reflect on how this responsibility can be ingrained in designs and reinforced through interactions.

Efficiency and accuracy

In the early design stages, some teachers believed *Pair-Up* could improve teacher efficiency when pairing students, while others were still determining its usefulness. Teachers across design activities were more favourable for shared control when *Pair-Up* provided accurate information while enabling teachers to leverage their own knowledge (Holstein et al., 2020). Accuracy and efficiency are two principles that designers need to consider regarding how teachers can leverage and act on information from AI-based technology (Shneiderman, 2022).

Reflections for HCLA practice

The focus of our paper was on the teacher's conceptualisation of shared control with *Pair-Up* throughout an iterative design process. In reporting our findings, we also illustrated how human centred processes can be leveraged to design human-AI technologies. While there were nuanced differences in the insights that emerged across design methods, we found a minimal change in teacher thoughts on control during dynamic pairing. Examining themes across all methods, we were able to see promising triangulation toward making design decisions about shared control based on teacher insights. Methodologically, it is challenging to attribute specific contributions to particular methods, especially when methods build on each other over time. Each design method had unique affordances that helped us to unpack different components of the dynamic pairing process and how AI might support teachers' instructional decisions. Our process built on examples of past processes (ie, Holstein et al., 2020), beginning with conceptualizing the root challenge of dynamic pairing and adding complexity over time. Starting in the classroom with a Wizard-of-Oz approach helped us to not only conceptualise the process but also unpack challenges of having the AI automate some work behind the scenes, which we were able to build upon over time.

We adapted traditional HCI and HCLA methods to engage teachers in designing our AI co-orchestration tool. One distinct difference for our HCAI process was the challenges of illustrating how pairing tasks might be automated with *Pair-Up*. Simulating ways to both explain and illustrate how the AI system might analyse, make predictions, and advise the dynamic pairing process was challenging through this process. Some teachers had a hard time imagining what the AI might do to support them and a few teachers felt that *Pair-Up* might do to support them and a few teachers felt that *Pair-Up* might be able to replace some of their usual interactions to alleviate some of the workload. Across all of our methods, the use of established HCLA methods that added simulated or imaginative prompts about the AI's automation and recommendations helped to support teachers in this process and allowed us to dig deeper into misunderstanding about what the AI could do or what its role in the classroom might be. In reflection of our process, we recommend methods that situate dynamic Human-AI interactions in authentic classroom practices to allow teachers to interact with prototypes with real data. Research shows this sheds light on more and better adoption (Xu et al., 2023). Additionally, simulating the classroom experience through real time replay of historic data has also proven to be an effective prototyping method, where users start to experience some of the new dynamics that the tool may bring about when used in the real environment (Holstein et al., 2019). HCLA are always grounded in practice to some extent, but our classroom and Wizard-of-Oz studies allowed teachers to contribute to the design in context to reflect on the use of AI and potential futures in their classrooms. Due to the COVID-19 pandemic, we could not work in classrooms as expected, but we found ways to ground design ideas in their context (eg, using real student names). Still, we must find ways to adapt methods to be more heavily grounded in teacher classroom norms and student learning needs to generate and evaluate design ideas.

In our last classroom activity, while pairing recommendations in *Pair-Up* were used, some teachers bypassed these to explore their manual pairings. It is hard to know if this was due to the novelty or usability of the tool or algorithmic aversion, or if this occurred due to contextually specific factors of their classroom. While examining our process across design methods helped understand teacher ideas about shared control, we suggest longitudinal methods to understand how teacher use of *Pair-Up* may change with sustained use (Xu et al., 2023). With a longer use case of the co-orchestration tool, we can better understand how teacher interactions and collaboration with *Pair-Up* sustain or change over time. With this in mind, we recognise that teachers should not simply accept *Pair-Up* suggestions, reinforcing the importance of *sharing* control.

Conclusions

Our research had two purposes. First, we described a secondary analysis to understand how teachers conceptualised sharing control with *Pair-Up*. Co-orchestration is one approach to leverage the strengths of teachers and AI to inform complex learning processes like dynamic transitions. However, AI co-orchestration tools must be designed with teachers to ensure the tools meet teacher needs and support student learning. Our findings show that teachers want to retain control when working with AI to co-orchestrate dynamic transitions. They want *Pair-Up* to supply suggestions and data, but teachers wish to retain the ultimate say over pedagogical decisions related to dynamic pairing (Molenaar, 2022). Additionally, teachers wanted *Pair-Up* to be accurate, flexible and customisable to support their classroom decisions (Shneiderman, 2022).

Second, we used our results to illustrate how HCLA can be applied to the design of human-AI technologies. Researchers suggest that HCI and HCLA methods require adaptation to account for the additional complexity of AI (Holstein et al., 2019; Ozmen Garibay et al., 2023; Xu et al., 2023). Our methods concretely addressed the Human-AI interactions within *Pair-Up*, including scaled-up storyboarding in a survey format, prompts to imagine the AI's role and Wizard-of-Oz studies. Our paper describes an entire process to create and test an AI-based co-orchestration tool with implications for teacher needs regarding sharing control with the AI. From this work, we see open areas to continue describing and innovating on how we adapt methods with teachers to build more collaborative human-AI technologies.

FUNDING INFORMATION

This work was funded by the National Science Foundation, grant number 1822861.

CONFLICT OF INTEREST STATEMENT

The authors do not have any interests that might be interpreted as influencing the research.

DATA AVAILABILITY STATEMENT

This work is a secondary analysis of previous published data, however, the work in the manuscript consists of original analysis and contributions.

ETHICS STATEMENT

Ethical standards were followed in the conduct of the study, including Institutional Review Board approval for all research activities and consent of all participants.

ORCID

LuEttaMae Lawrence https://orcid.org/0000-0001-6066-5096
Vanessa Echeverria https://orcid.org/0000-0002-2022-9588

REFERENCES

Buchenau, M., & Suri, J. F. (2000, August). Experience prototyping. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 424–433).

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, *6*(2), 1–9.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*, 253–278.

Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In *New science of learning: Cognition, computers and collaboration in education* (pp. 525–552). Springer.

Dimitriadis, Y., Martínez-Maldonado, R., & Wiley, K. (2021). Human-centered design principles for actionable learning analytics. In *Research on e-learning and ICT in education: Technological, pedagogical and instructional perspectives* (pp. 277–296). Springer.

Dubey, A., Abhinav, K., Jain, S., Arora, V., & Puttaveerana, A. (2020). HACO: A framework for developing human-AI teaming. In *Proceedings of the 13th Innovations in Software Engineering Conference* (pp. 1–9).

Echeverria, V., Yang, K., Lawrence, L., Rummel, N., & Aleven, V. (2023). Designing hybrid human–AI orchestration tools for individual and collaborative activities: A technology probe study. *IEEE Transactions on Learning Technologies*, *16*(2), 191–205.

Fang, Y., Roscoe, R. D., & McNamara, D. S. (2023). Artificial intelligence-based assessment in education. In B. du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of artificial intelligence in education* (pp. 485–504). Edward Elgar Publishing.

Goodyear, P., & Dimitriadis, Y. (2013). Reframing design for learning. *Research in Learning Technology*, *21*, 19909. https://doi.org/10.3402/rlt.v21i0.19909

Haesen, M., Meskens, J., Luyten, K., & Coninx, K. (2010). Draw me a storyboard: Incorporating principles & techniques of comics. *Proceedings of HCI 2010*, *24*, 133–142.

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In *Artificial Intelligence in Education, Ifrane, Morocco*, July 6–10, 2020, Proceedings, Part I (pp. 240–254). Springer International Publishing.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real time classroom orchestration tool to support teacher-AI complementarity. *Journal of Learning Analytics*, *6*(2), 27–52.

Holstein, K., & Olsen, J. K. (2023). 14. Human–AI co-orchestration: The role of artificial intelligence in orchestration. In *Handbook of artificial intelligence in education* (pp. 309–321). Edward Elgar Publishing.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., & Eiderbäck, B. (2003). Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 17–24).

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *5*(36), 757–798.

Lawrence, L., Guo, B., Yang, K., Echeverria, V., Kang, Z., Bathala, V., Li, C., Huang, W., Rummel, N., & Aleven, V. (2022). Process to co-design AI-based orchestration tools to support dynamic transitions: Design narratives through Conjecture Mapping. In A. Weinberger, W. Chen, D. Hernández-Leo, & B. Chen (Eds.), *Proceedings of the 15th international conference on computer-supported collaborative learning-CSCL 2022* (pp. 139–146). International Society of the Learning Sciences.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. Springer.

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, *57*(4), 632–645.

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning*, *6*(3), 421–443.

Nielsen, J., Clemmensen, T., & Yssing, C. (2002, October). Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction* (pp. 101–110).

Olsen, J. K., Aleven, V., & Rummel, N. (2015). Toward combining individual and collaborative learning within an intelligent tutoring system. In *Artificial intelligence in education: 17th international conference*, AIED 2015, Madrid, Spain, June 22-26, 2015. Proceedings 17 (pp. 848–851). Springer International Publishing.

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., … Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391–437.

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, *24*, 33–61.

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. *International Journal of Human–Computer Interaction*, *39*(3), 494–518.

Yang, K. B., Echeverria, V., Lu, Z., Mao, H., Holstein, K., Rummel, N., & Aleven, V. (2023). Pair-up: prototyping human-AI co-orchestration of dynamic transitions between individual and collaborative learning in the classroom. In S. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–17). Association for Computing Machinery.

Yang, K. B., Echeverria, V., Wang, X., Lawrence, L., Holstein, K., Rummel, N., & Aleven, V. (2021). Exploring Policies for Dynamically Teaming up Students through Log Data Simulation. In I. Hsiao, S. Sahebi, F. Bouchet, & J. J. Vie (Eds.), *Proceedings, fourth international educational data mining society* (pp. 183–194). Educational Data Mining Society.

Yang, K. B., Lawrence, L., Echeverria, V., Guo, B., Rummel, N., & Aleven, V. (2021). Surveying teachers' preferences and boundaries regarding human-AI control in dynamic pairing of students for collaborative learning. In *Technology-enhanced learning for a free, safe, and sustainable world: 16th european conference on technology enhanced learning*, EC-TEL 2021, Bolzano, Italy, September 20-24, 2021, Proceedings 16 (pp. 260–274). Springer International Publishing.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020, April). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).

How to cite this article: Lawrence, L., Echeverria, V., Yang, K., Aleven, V., & Rummel, N. (2023). How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology*, *00*, 1–22. https://doi.org/10.1111/bjet.13372