Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms

Kenneth Holstein(✉), Bruce M. McLaren, and Vincent Aleven

Carnegie Mellon University, Pittsburgh, PA 15213, USA
{kjholste,bmclaren,aleven}@cs.cmu.edu

Abstract. As artificial intelligence (AI) increasingly enters K-12 classrooms, what do teachers and students see as the roles of human versus AI instruction, and how might educational AI (AIED) systems best be designed to support these complementary roles? We explore these questions through participatory design and needs validation studies with K-12 teachers and students. Using human-centered design methods rarely employed in AIED research, this work builds on prior findings to contribute: (1) an analysis of teacher and student feedback on 24 design concepts for systems that integrate human and AI instruction; and (2) participatory speed dating (PSD): a new variant of the speed dating design method, involving iterative concept generation and evaluation with multiple stakeholders. Using PSD, we found that teachers desire greater real-time support from AI tutors in identifying when students need human help, in evaluating the impacts of their own help-giving, and in managing student motivation. Meanwhile, students desire better mechanisms to signal help-need during class without losing face to peers, to receive emotional support from human rather than AI tutors, and to have greater agency over how their personal analytics are used. This work provides tools and insights to guide the design of more effective human–AI partnerships for K-12 education.

Keywords: Design · Classroom orchestration · Human-AI interaction

1 Introduction

When used in K-12 classrooms, AI tutoring systems (ITSs) can be highly effective in helping students learn (e.g., [32,37]). However, in many situations, human teachers may be better suited to support students than automated systems alone (e.g., by providing socio-emotional support or flexibly providing conceptual support when continued problem-solving practice may be insufficient) [29,44,49,53]. ITSs might be even more effective if they were designed not only to support students directly, but also to take advantage of teachers' complementary strengths and amplify their abilities to help their students [6,27,49,65]. Yet the question

S. Isotani et al. (Eds.): AIED 2019, LNAI 11625, pp. 157–171, 2019.
https://doi.org/10.1007/978-3-030-23204-7_14

of how best to combine strengths of human and AI instruction has received relatively little attention in the AIED literature thus far [29,49,60].

Recent work has proposed the notion of human–AI "co-orchestration" systems that help teachers and AI agents work together to make complex yet powerful learning scenarios feasible [27,29,43,46,54,60]. For example, Olsen et al. explored how ITSs might best be designed to share control with teachers in orchestrating transitions between individual and collaborative activities during a class session [15,43]. Similarly, in our prior work [26,28,29], we designed a set of mixed-reality smart glasses that direct teachers' attention in real-time, during ITS class sessions, towards situations the software may be ill-suited to handle on its own (e.g., wheel spinning [7,31], gaming the system [5,58], or hint avoidance [2,51]). An in-vivo classroom experiment demonstrated that this form of real-time teacher/AI co-orchestration could enhance student learning, compared with an ITS classroom in which the teacher did not have such support [29].

While this work has begun to explore ways to combine strengths of human and AI instruction, many open questions remain regarding the design of classroom co-orchestration systems. If these tools are to be used in actual classrooms, beyond the context of research studies, it is critical that they are well-designed to respect the needs and boundaries of both teachers and students [3,14,42,52,66]. For example, prior design research with K-12 teachers has found that there is a delicate balance between automation and respecting teachers' autonomy [25,27,34,43]. Over-automation may take over classroom roles that teachers would prefer to perform and threaten their flexibility to set their own instructional goals. Yet under-automation may burden teachers with tasks they would rather not perform, and may limit the degree of personalization they can feasibly achieve in the classroom [27,43]. Furthermore, this balance may depend heavily on the specific teacher tasks under consideration [26,55]. Yet prior work on co-orchestration systems has investigated the design of support for a relatively limited range of teacher tasks (e.g., monitoring student activities during class [45,50]). Furthermore, this research has generally focused on the needs of K-12 teachers, but not students' perspectives, in AI-enhanced classrooms [27,34,43].

The present work builds on prior findings to contribute: (1) an analysis of teacher and student feedback regarding 24 design concepts for human–AI co-orchestration systems, to understand key needs and social boundaries that such systems should be designed to address [13,21,66] and (2) "participatory speed dating": a new variant of the speed dating design method [12] that involves multiple stakeholders in the generation and evaluation of novel technology concepts.

2 Methods

To better understand and validate needs uncovered in prior ethnographic and design research with K-12 students and teachers (e.g., [20,27,43,52,53]), we adopted a participatory speed dating approach. Speed dating is an HCI method for rapidly exploring a wide range of possible futures with users, intended to help researchers/designers elicit unmet needs and probe the boundaries of what particular user populations will find acceptable (which otherwise often remain undiscovered until after a technology prototype has been developed and deployed) [12,42,67]. In speed dating sessions, participants are presented with a number of hypothetical scenarios in rapid succession (e.g., via storyboards) while researchers observe and aim to understand participants' immediate reactions.

Speed dating can lead to the discovery of unexpected design opportunities, when unanticipated needs are uncovered or when anticipated boundaries are discovered not to exist. Importantly, speed dating can often reveal needs and opportunities that may not be observed through field observations or other design activities [12,13,42,67]. For example, Davidoff et al. found that, whereas field observations and interview studies with parents had suggested they might appreciate smart home technologies that automate daily household tasks, a speed dating study revealed that parents strongly rejected the idea of automating certain tasks, such as waking or dressing their children in the morning. These findings led the researchers to dramatically reframe their project—away from creating smart homes that "do people's chores," towards homes that facilitate moments of bonding and connection between busy family members [12,67].

As described in the next subsection, we adapted the speed dating method to enable participants from multiple stakeholder groups (K-12 teachers and students) to reflect on other stakeholders' needs and boundaries, and contribute ideas for new scenarios and technology concepts. We refer to this adaptation as multi–stakeholder "participatory speed dating" (PSD). Like other speed dating approaches, PSD can help to bridge between broad, exploratory design phases and more focused prototyping phases (where associated costs may discourage testing a wide range of ideas) [12,18,67]. However, drawing from Value Sensitive Design [21,66], PSD emphasizes a systematic approach to balancing multiple stakeholder needs and values [38]. Drawing from Participatory Design [36,40,56], in addition to having stakeholders evaluate what is wrong with a proposed concept (which may address other stakeholders' needs), PSD also involves them in generating alternative designs, to address conflicts among stakeholder groups.

2.1 Needs Validation Through Participatory Speed Dating

We conducted PSD sessions one-on-one with 24 middle school teachers and students. To recruit participants, we emailed contacts at eight middle schools and advertised the study on Nextdoor, Craigslist, and through physical fliers. A total of 10 teachers and 14 students, from two large US cities, participated in the study. Sixteen sessions were conducted face-to-face at our institution, and eight were conducted via video conferencing. All participants had experience using some form of adaptive learning software in their classrooms, and 21 participants had used AI tutoring software such as ALEKS [23] or Cognitive Tutor [48].

We first conducted a series of four 30-minute study sessions focused on concept generation, with two teachers and two students. In each session, participants were first introduced to the context for which they would be designing: classes in which students work with AI tutoring software while their teacher uses a real-time co-orchestration tool that helps them help their students (specifically, a set of teacher smart glasses, following [29]). Participants were then shown an initial set of 11 storyboards, each created to illustrate specific classroom challenges uncovered in prior research (e.g., [20,27,47,53]), with multiple challenges hybridized [12,42] into a single storyboard in some cases.[1] For example, prior work suggests that teachers often struggle to balance their desire to implement personalized, mastery-based curricula with their need to keep the class relatively synchronized and "on schedule" [27]. Given this conflict, teachers often opt to manually push students forward in the curriculum if they have failed to master current skills in the ITS by a certain date, despite awareness that this practice may be harmful to students' learning [27,47]. As such, one storyboard (Fig. 1) presented a system that helps teachers make more informed decisions about when to move students ahead (based on the predicted learning benefits of waiting a few more class periods), but without strongly suggesting a particular course of action [27]. Each participant in these initial studies was then encouraged to generate at least one new idea for a storyboard, addressing challenges they personally face in AI-enhanced classrooms as opposed to imagined challenges of others (cf. [13]). To inform ideation, participants also reviewed storyboards generated by other teachers and students in prior study sessions. Participants were provided with editable storyboard templates, in Google Slides [22], and were given the options to generate entirely new concepts for orchestration tool functionality (starting from a blank template) or to generate a variation on an existing concept (starting from a copy of an existing storyboard). In either case, participants generated captions for storyboard panels during the study session, using existing storyboards for reference. Immediately following each session, a researcher then created simple illustrations to accompany each caption.

Following this concept generation phase, we conducted a series of PSD studies with an additional twelve students and eight teachers. Study sessions lasted approximately 60 min. In each session, storyboards were presented in randomized order. Participants were asked to read each storyboard and to describe their initial reactions immediately after reading each one. An interviewer asked follow-up and clarification questions as needed. Participants were then asked to provide an overall summary rating of the depicted technology concept as "mostly positive (I would probably want this feature in my classroom)", "mostly negative (I would probably not want this ...)", or "neutral" [13]. After participants rated each concept, they were asked to elaborate on their reasons for this rating. Before moving on to the next concept, participants were shown notes on reactions to a given concept, thus far, from other stakeholders. Participants were prompted to share their thoughts on perspectives in conflict with their own.

In addition, participants were encouraged to pause the speed dating process at any point, if they felt inspired to write down an idea for a new storyboard. Each time a participant generated a new idea for a storyboard, this storyboard was included in the set shown to the next participant. However, if a participant

[1] Please refer to https://tinyurl.com/Complementarity-Supplement for the full set of storyboards and more detailed participant demographics.

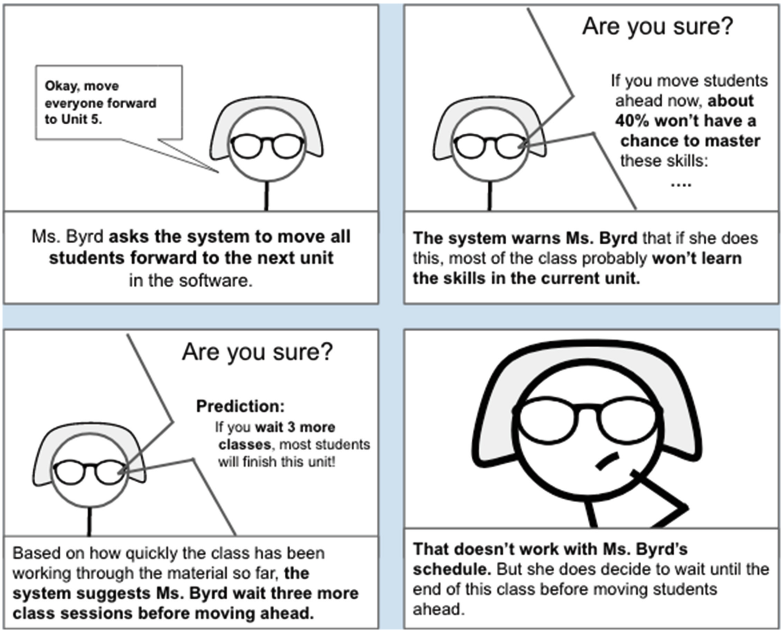

Fig. 1. Example of a storyboard addressing challenges raised in prior research.

saw an existing storyboard that they felt captured the same concept as one they had generated, the new, "duplicate" storyboard was not shown to subsequent participants (cf. [27]). In cases of disagreement between stakeholder groups, generating new storyboard ideas provided an opportunity for students and teachers to try to resolve these disagreements. For example, as shown in Fig. 2, the generation of concepts E.3 through E.6 over time represents a kind of "negotiation" between teachers and students, around issues of student privacy, transparency, and control. This phase of the study yielded a total of seven new storyboards.

3 Results

In the following subsections, we discuss teachers' and students' top five most and least preferred design concepts, according to the average overall ratings among those who saw a given concept [13]. To analyze participant feedback regarding each concept, we worked through transcriptions of approximately 19 h of audio to synthesize findings using two standard methods from Contextual Design: interpretation sessions and affinity diagramming [8,24]. High-level themes that emerged are briefly summarized below, organized by design concept. The most preferred concepts are presented in Sect. 3.1, and the least preferred are in Sect. 3.2. Within each subsection, preferences among teachers are presented first, followed by student preferences and those shared between teachers and students. Teacher participants are identified with a "T," and students are identified with an "S."

	S1	S2	S3	S4	T1	T2	T3	T4	S5	S6	S7	T5	S8	T6	T7	S9	T8	S10	S11	S12	Teacher avg.	Student avg.
[A.1] Ranking Students by Need for Teacher Help	0	1	0	1	1	1	0	1	1	-1	1	1	1	1	1	0	1	0	1	1	0.88	0.50
[A.2] Explaining Ranking of Students	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	1	0	0	1	0.13	0.08
[B] Suggesting Which Students to Help and How to Help	0	1	0	0	1	1	-1	1	1	0	0	1	0	1	1	0	1	0	0	1	0.75	0.25
[C] Helping Teachers Mediate between Stu. and Student Models	0	1	1	-1	-1	0	1	-1	1	0	0	-1	1	1	-1	1	-1	-1	0	1	-0.38	0.33
[D] Predicting Time to Mastery to Support Teacher Scheduling	0	1	0	0	1	1	0	1	1	0	1	1	0	-1	1	0	1	0	0	1	0.63	0.33
[E.1] Alerting Teachers to Student Frustration, Misbehavior, ...	0	-1	-1	0	1	1	0	1	1	0	1	1	-1	1	1	1	1	1	-1	1	0.88	0.08
[E.2] Providing Automated Motivational Prompts ...	-1	-1	-1	0	-1	1	1	-1	1	0	1	-1	-1	0	0	1	0	1	-1	1	-0.13	0.00
[E.3] Allowing Stu. to Hide (All) of their Analytics from Teachers			1	-1	-1	-1	-1	-1	0	0	-1	1	1	-1	-1	-1	-1	-1	0	-1	-0.75	-0.30
[E.4] Notifying Stu. When the System has Alerted their Teacher							1	-1	1	0	1	1	0	1	0	0	0	-1	-1	-1	0.33	-0.13
[E.5] Allowing Students to Hide Emotion-related Analytics ...													1	0	-1	0	1	0	1	1	0.00	0.60
[E.6] Asking Stu. Permission before Revealing (Some) Analytics ...														0	0	1	0	1	1	1	0.00	1.00
[F.1] "Invisible Hand Raises" and Teacher Reminders	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	-1	0	1	0.88	0.75
[F.2] Suggesting Peer Tutors to Support Teachers ...	1	1	-1	-1	1	1	1	1	0	1	1	1	-1	1	0	0	0	1	0	1	0.75	0.25
[G] Providing Teacher with Suggested "Conversation Starters" ...	0	1	0	0	-1	1	1	1	1	1	0	1	0	1	-1	0	1	0	0	1	0.50	0.33
[H.1] Enabling Students to Request Not to be Helped	0	1	1	1	1	1	0	-1	1	0	1	-1	1	1	1	1	1	1	1	1	0.38	0.83
[H.2] Enabling Stu. to Ask the Whole Class Anonymous Questions	1	1	1	0	1	0	0	0	1	0	1	-1	1	1	-1	1	1	1	0	0	0.13	0.67
[H.3] Student–System Joint Control Over Selection of Peer Tutors				1	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1	1	0.63	0.89
[H.4] Showing Students Potential Peer Tutors' Skill Mastery											1	1	-1	-1	-1	-1	0	0	-1	-1	-0.25	-0.50
[I.1] Real-time Positive Feedback on Teacher Explanations.	1	1	1	0	0	1	0	1	0	1	0	1	1	1	1	0	1	1	0	1	0.75	0.58
[I.2] Real-time Negative Feedback on Teacher Explanations.	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	0	1	1	0	1	1.00	0.58
[J] Notifying Teachers about Stu. they Have Not Visited Recently	0	1	1	1	1	1	0	-1	1	0	1	1	1	1	-1	1	1	1	1	1	0.38	0.83
[K] Listening in on Teacher Help-giving to Improve AI Tutor's Hints	0	0	-1	-1	-1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0.75	0.25
[L] Teacher-controlled Shared Displays to Foster Competition	1	1	1	1	1	1	1	1	0	1	-1	0	1	1	1	0	1	1	1	1	0.88	0.67
[M] Allowing Parents to Monitor their Child's Behavior During Class							1	-1	0	0	0	-1	-1	1	-1	-1	0	-1	-1	0	-0.17	-0.50

Fig. 2. Matrix showing overall ratings for all 24 concepts. Columns show participants (in order of participation, from left to right), and rows show design concepts. Concepts generated by participants are highlighted in blue. Cell colors indicate ratings as follows: Red: negative; Green: positive; Yellow: neutral; Grey: concept did not yet exist. Average ratings among teachers and students are provided in the rightmost columns. (Color figure online)

3.1 Most Preferred Design Concepts

Most Preferred Among Teachers

[I.2] Real-time Feedback on Teacher Explanations. Consistent with findings from prior design research [26,27], the most popular concept among teachers was a system that would provide them with constructive feedback, after helping a student, on the effectiveness of their own explanations. As one teacher explained, *"Usually our only chance to get [fast] feedback is, you ask [...] the kids [and] they just say, 'Oh, yeah, I get it,' when they don't really get it"* (T7).

[A.1] Ranking Students by their Need for Teacher Help. Another popular concept among teachers was a system that would allow them to see, at a glance, a visual ranking of which students most need the teacher's help at a given moment [27,49]. One teacher commented, *"Yeah. Welcome to teaching every day [...] trying to go to those kids that are [struggling] most"* (T5). However, several other teachers emphasized that such a ranking would be much more useful if it took into account the kind and extent of teacher help that would likely be needed to address a particular student issue. For example, T1 noted, *"If I could see how much time it would take [to help] I would start with the kids who I could get [moving again quickly] and then I'd spend more time with the other kids. [But] if it's a kid that I know is gonna get completely frustrated [...then I] wanna [go to] that kid first no matter what."* This concept was also generally well received by students. As one student put it, *"sometimes you just can't ask [for help] because you don't even know what [you're struggling with], and so it would just [be] hard to explain it to the teacher"* (S7). At the same time, as discussed

below, multiple students expressed preferences for systems that can support *students* in recognizing when (and with what) they need to ask the teacher for help, rather than always having the system alert the teacher on their behalf (cf. [51]).

[E.1] Alerting Teachers to Student Frustration, Misbehavior, or "Streaks". Consistent with [27], teachers were enthusiastic about a concept that would allow them to see real-time analytics about student frustration, misbehavior (e.g,. off-task behavior or gaming the system [5,58]), or high recent performance in the software. They felt that having access to this information could help them make more informed decisions about whom to help first and how best to help particular students (e.g., comforting a student or offering praise). Yet students reported finding aspects of this concept upsetting. While students generally liked the idea that the system would inform the teacher when they needed help, students often perceived real-time alerts about emotions like frustration as *"really creepy"* (S9) and alerts about misbehavior as *"basically the AI ratting out the child"* (S3).

[L] Teacher-controlled Shared Displays to Foster Competition. Finally, a popular concept among teachers was a system that would allow them to transition the classroom between different "modes," to help regulate students' motivation (cf. [1,43]). This system would allow teachers to switch the class into a "competitive mode," in which students would be shown a leaderboard of comparable classrooms in their school district and challenged to move their class to the top. Teachers expected that such a feature could work extremely well with some groups of students, while backfiring and potentially serving to demotivate others. As such, teachers emphasized the importance of teacher control and discretion.

Most Preferred Among Students

[E.6] Asking Students' Permission before Revealing (Some) Analytics to Teachers. In response to one of teachers' most preferred design concepts (*[E.1]*), students generated multiple new storyboards that preserved the idea of real-time teacher alerts, but provided students with greater control over alert policies. One of these emerged as the most popular design concept among students: a system that asks students' permission, on a case-by-case basis, before presenting certain kinds of information to the teacher on a student's behalf. Students and teachers were generally in agreement that an AI system should ask students' permission before alerting teachers about affective states, such as frustration. In this scenario, if a student opted not to share affective analytics with their teacher, the system might privately suggest other ways for students to regulate their own emotions. Interestingly, one student suggested that if a student opted to share their affect with the teacher, the system should also ask the student to specify *"How do you want the teacher to react? [...] Help you [in person]? Help you on the computer?"* (S12). This student noted that sometimes, they just want their teacher to *"know how I'm feeling,"* but do not actually want them to take action.

[H.3] Student–System Joint Control Over Selection of Peer Tutors. Whereas teachers often expressed that they know which groups of their students will not work well together, this did not align with students' perceptions of their own teachers. In contrast to teacher-generated concepts where teachers and AI worked together to match peer tutors and tutees (cf. [43]), the second most popular

concept among students was a student-generated storyboard that gave students the final say over peer matching decisions. In this storyboard, the system sends struggling students a list of suggested peer tutors, based on these students' estimated tutoring abilities (cf. [57]) and knowledge of relevant skills. Students could then send help requests to a subset of peers from this list who they would feel comfortable working with. Those invited would then have the option to reject a certain number of requests. Some students suggested that it would also be useful to have the option to accept but delay another student's invitation if they want to help but do not want to disrupt their current flow.

[H.1] Enabling Students to Request Not to be Helped. Another of the most popular concepts among students was a system that, upon detecting that a student seems to be wheel-spinning [7,31], would notify the student to suggest that they try asking their teacher or classmates for help. The system would then only notify the teacher that the student is struggling if the student both ignored this suggestion and remained stuck after a few minutes. By contrast, some teachers expressed that they would want the system to inform them *immediately* if a student was wheel-spinning: *"They shouldn't just get the option to keep working on their own, because honestly it hasn't been working"* (T5). Some students and teachers suggested a compromise: *"the AI should inform the teacher right away [...] that it suggested [asking for help] but the kid did something else"* (T7).

[J] Notifying Teachers of Students they Have Not Visited Recently. Finally, a popular concept among students was a system that would track a teachers' movement during class and occasionally highlight students they may be neglecting (cf. [4,19]). Several students noted that even when they are doing well on their own, they feel motivated when their teacher remembers to check in with them. Most teachers responded positively to this concept as, *"sometimes you forget about the kids that work well on their own, but sometimes those kids actually need help and don't raise their hands"* (T6). However, a few teachers perceived this system as overstepping bounds and inappropriately judging them: *"It's just too much in my business now. You better be quiet and give me a break"* (T4).

Most Preferred Among Both Teachers and Students

[F.1] "Invisible Hand Raises" and Teacher Reminders. A concept popular with both teachers and students was a system that would allow students to privately request help from their teacher by triggering an "invisible hand raise" that only the teacher could see. To preserve privacy, this system would also allow teachers to silently acknowledge receipt of a help request. After a few minutes, the teacher would receive a light reminder if they had not yet helped a student in their queue. S7 noted, *"I don't actually like asking questions since I'm supposed to be, like, 'the smart one' ...which I'm not. So I like the idea of being able to ask a question without [letting] others know."* Similarly, teachers suspected that students would request help more often if they had access to such a feature [26,53].

3.2 Least Preferred Design Concepts

Least Preferred Among Teachers

[C] Helping Teachers Mediate between Students and their Student Models. To our surprise, although prior field research [30] had suggested teachers might find it desirable to serve as "final judges" in cases where students wished to contest their student models (e.g., skill mastery estimates) [11], this was one of the least popular design concepts among teachers. Students generally viewed teacher-in-the-loop mediation desirable, since *"I feel like the teacher knows the student better, not the software"* (S9). However, teachers generally did not view this as an efficient use of their time: *"I would just trust the tutor on this one"* (T3). Furthermore, some teachers expressed concerns that from a student's perspective this concept *"pit[s] one teacher against the other, if you consider the AI as a kind of teacher"* (T1), and instead suggested having the system assign a targeted quiz if a student wants to demonstrate knowledge of particular skills (cf. [11]).

Least Preferred Among Students

[E.4] Notifying Students When the System has Automatically Alerted their Teacher. A teacher-generated concept intended to provide students with greater transparency into the analytics being shared about them was among those least popular with students overall. Interestingly, while students valued having more control over the information visible to their teachers, they generally did not want greater transparency into aspects of the system that were outside of their control (cf. [33]): *"That would make me really anxious [...] If it's not asking students' [permission], I don't think they should know about it"* (S10).

Least Preferred Among Both Teachers and Students

[E.3] Allowing Students to Hide (All) of their Analytics from Teachers. The least popular concept among teachers, and the third least popular among students, was a privacy feature that would enable individual students to prevent their AI tutor from sharing real-time analytics with their teacher. This was a student-generated concept intended to mitigate the "creepiness" of having their teacher "surveil" students' activities in real-time. Yet as discussed in Sect. 3.1, overall students felt that it should only be possible for students to hide certain kinds of analytics (e.g., inferred emotional states), *"but if the AI sees a student is really, really struggling [...] I don't think there should be that blanket option"* (S4).

[H.4] Showing Students Potential Peer Tutors' Skill Mastery. Consistent with prior research (e.g., [26]), teachers and students responded negatively to a student-generated concept that made individual students' skill mastery visible to peers. While this concept was intended to help students make informed choices about whom to request as a peer tutor, most teachers and students perceived that the risk of teasing among students outweighed the potential benefits.

[M] Allowing Parents to Monitor their Child's Behavior During Class. Somewhat surprisingly, T3 generated the concept of a remote monitoring system that would allow parents to *"see exactly what [their child is] doing at any moment in time."*, so that *"if a kid's misbehaving, their parent can see the teacher's trying [their] best"* (cf. [9,62]). While this concept resonated with one other teacher,

student and teacher feedback on this concept generally revealed an attitude that to create a safe classroom environment, *"we have to [be able to] trust that data from the classroom stays in the classroom"* (S11). Teachers shared concerns that data from their classrooms might be interpreted out of context by administrators: *"I don't ever want to be judged as a teacher [because] I couldn't make it to every student, if every kid's stuck that day. [But] using that data [as a teacher] is very useful"* (T5). Students shared fears that, depending on the data shared, parents or even future employers might use classroom data against them.

[E.2] Providing Automated Motivational Prompts to Frustrated Students. Finally, among the concepts least popular with both teachers and students was a system that automatically provides students with motivational prompts when it detects they are getting frustrated [16,64]. While teachers liked the idea of incorporating gamification elements to motivate students (cf. [35,62]), providing motivational messages was perceived as *"trying to [do] the teacher's job"* (T1). Similarly, several students indicated strongly that they would prefer these kinds of messages to come from an actual person, if at all. S8 said, *"I would just get more annoyed if the AI tried something like that"*, and S11 suggested *"No emotional responses, please. That feels just [...] not genuine. If it's from the AI it should be more analytical, like just [stick to] facts."*

4 Discussion, Conclusions, and Future Work

If new AI systems are to be well-received in K-12 classrooms, it is critical that they support the needs and respect the boundaries of both teachers and students. We have introduced "participatory speed dating" (PSD): a variant of the speed dating design method that involves multiple stakeholders in the iterative generation and evaluation of new technology concepts. Using PSD, we sampled student and teacher feedback on 24 design concepts for systems that integrate human and AI instruction—an important but underexplored area of AIED research.

Overall, we found that teachers and students aligned on needs for "hidden" student–teacher communication channels during class, which enable students to signal help-need or other sensitive information without losing face to their peers. More broadly, both teachers and students expressed nuanced needs for student privacy in the classroom, where it is possible to have "too little," "too much," or the wrong forms of privacy (cf. [41]). However, students and teachers did not always perceive the same needs. As discussed in Sect. 3.1, some of students' highest-rated concepts related to privacy and control were unpopular among teachers. Additional disagreements arose when teachers and students had different expectations of the roles of teachers versus AI agents and peer tutors in the classroom.

Interestingly, while students' expressed desires for transparency, privacy, and control over classroom AI systems extend beyond what is provided by existing systems [9,11,29,60], these desires are also more nuanced than commonly captured in theoretical work [10,59,61]. For example, we found that while students were uncomfortable with AI systems sharing certain kinds of personal analytics with their teacher without permission, they rejected design concepts that grant students full control over these systems' sharing policies. These findings indicate an important role for empirical, design research approaches to complement critical and policy-oriented research on AI in education (cf. [33,41,63]).

In sum, the present work provides tools and and early insights to guide the design of more effective and desirable human–AI partnerships for K-12 education. Future AIED research should further investigate teacher and student needs uncovered in the present work via rapid prototyping in live K-12 classrooms. While design methods such as PSD are critical in guiding the initial development of novel prototypes, many important insights surface only through deployment of functional systems in actual, social classroom contexts [30,42,53]. An exciting challenge for future research is to develop methods that extend the advantages of participatory and value-sensitive design approaches (e.g., [39,56,66]) to later stages of the AIED design cycle. Given the complexity of data-driven AI systems [17,26,66], fundamentally new kinds of design and prototyping methods may be needed to enable non-technical stakeholders to remain meaningfully involved in shaping such systems, even as prototypes achieve higher fidelity.

Acknowledgements. This work was supported by IES Grants R305A180301 and R305B150008. The opinions expressed are those of the authors and do not represent the views of IES or the U.S. ED. Special thanks to all participating teachers and students.

References

1. Alavi, H.S., Dillenbourg, P.: An ambient awareness tool for supporting supervised collaborative problem solving. IEEE Trans. Learn. Technol. **5**(3), 264–274 (2012)
2. Aleven, V., Roll, I., McLaren, B.M., Koedinger, K.R.: Help helps, but only so much: research on help seeking with intelligent tutoring systems. Int. J. Artif. Intell. Educ. **26**(1), 205–223 (2016)
3. Amershi, S., et al.: Guidelines for human-AI interaction. ACM, May 2019. https://www.microsoft.com/en-us/research/publication/guidelines-for-human-ai-interaction/
4. An, P., Bakker, S., Ordanovski, S., Taconis, R., Eggen, B.: Classbeacons: designing distributed visualization of teachers-physical proximity in the classroom. In: Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 357–367. ACM (2018)
5. Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K.: Why students engage in "gaming the system" behavior in interactive learning environments. J. Interact. Learn. Res. **19**(2), 185–224 (2008)
6. Baker, R.S.: Stupid tutoring systems, intelligent humans. Int. J. Artif. Intell. Educ. **26**(2), 600–614 (2016)
7. Beck, J.E., Gong, Y.: Wheel-spinning: students who fail to master a skill. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 431–440. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_44
8. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Elsevier, San Francisco (1997)

9. Broderick, Z., O'Connor, C., Mulcahy, C., Heffernan, N., Heffernan, C.: Increasing parent engagement in student learning using an intelligent tutoring system. J. Interact. Learn. Res. **22**(4), 523–550 (2011)
10. Bulger, M.: Personalized learning: the conversations we're not having. Data Soc. **22** (2016)
11. Bull, S., Kay, J.: Smili: a framework for interfaces to learning data in open learner models, learning analytics and related fields. Int. J. Artif. Intell. Educ. **26**(1), 293–331 (2016)
12. Davidoff, S., Lee, M.K., Dey, A.K., Zimmerman, J.: Rapidly exploring application design through speed dating. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 429–446. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74853-3_25
13. Dillahunt, T.R., Lam, J., Lu, A., Wheeler, E.: Designing future employment applications for underserved job seekers: a speed dating study. In: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 33–44. ACM (2018)
14. Dillenbourg, P.: Design for classroom orchestration. Comput. Educ. **69**, 485–492 (2013)
15. Dillenbourg, P., Prieto, L.P., Olsen, J.K.: Classroom orchestration (2018)
16. D'Mello, S., Picard, R.W., Graesser, A.: Toward an affect-sensitive autotutor. IEEE Intell. Syst. **22**(4), 53–61 (2007)
17. Dove, G., Halskov, K., Forlizzi, J., Zimmerman, J.: UX design innovation: challenges for working with machine learning as a design material. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 278–288. ACM (2017)
18. Dow, S.P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D.L., Klemmer, S.R.: Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM Trans. Comput. Hum. Interact. (TOCHI) **17**(4), 18 (2010)
19. Echeverria, V., Martinez-Maldonado, R., Power, T., Hayes, C., Shum, S.B.: Where is the nurse? towards automatically visualising meaningful team movement in healthcare education. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10948, pp. 74–78. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_14
20. Feng, M., Heffernan, N.T.: Informing teachers live about student learning: reporting in the assistment system. Technol. Instr. Cogn. Learn. **3**(1/2), 63 (2006)
21. Friedman, B.: Value-sensitive design. Interactions **3**(6), 16–23 (1996)
22. Google: Google slides (2019). http://slides.google.com
23. Hagerty, G., Smith, S.: Using the web-based interactive software aleks to enhance college algebra. Math. Comput. Educ. **39**(3) (2005)
24. Hanington, B., Martin, B.: Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions. Rockport Publishers, Beverly (2012)
25. Heer, J.: Agency plus automation: designing artificial intelligence into interactive systems. Proc. Nat. Acad. Sci. **116**(6), 1844–1850 (2019). https://doi.org/10.1073/pnas.1807184115, https://www.pnas.org/content/early/2019/01/29/1807184115
26. Holstein, K., Hong, G., Tegene, M., McLaren, B.M., Aleven, V.: The classroom as a dashboard: co-designing wearable cognitive augmentation for k-12 teachers. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 79–88. ACM (2018)

27. Holstein, K., McLaren, B.M., Aleven, V.: Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 257–266. ACM (2017)
28. Holstein, K., McLaren, B.M., Aleven, V.: Informing the design of teacher awareness tools through causal alignment analysis. In: International Conference of the Learning Sciences, pp. 104–111 (2018)
29. Holstein, K., McLaren, B.M., Aleven, V.: Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 154–168. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_12
30. Holstein, K., McLaren, B.M., Aleven, V.: Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. J. Learn. Anal. (Under review)
31. Kai, S., Almeda, M.V., Baker, R.S., Heffernan, C., Heffernan, N.: Decision tree modeling of wheel-spinning and productive persistence in skill builders. JEDM—J. Educ. Data Min. **10**(1), 36–71 (2018)
32. Kulik, J.A., Fletcher, J.: Effectiveness of intelligent tutoring systems: a meta-analytic review. Rev. Educ. Res. **86**(1), 42–78 (2016)
33. Lee, M.K., Baykal, S.: Algorithmic mediation in group decisions: fairness perceptions of algorithmically mediated vs. discussion-based social division. In: CSCW, pp. 1035–1048 (2017)
34. van Leeuwen, A., et al.: Orchestration tools for teachers in the context of individual and collaborative learning: what information do teachers need and what do they do with it? In: International Society of the Learning Sciences, Inc.[ISLS] (2018)
35. Long, Y., Aman, Z., Aleven, V.: Motivational design in an intelligent tutoring system that helps students make good task selection decisions. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M.F. (eds.) AIED 2015. LNCS (LNAI), vol. 9112, pp. 226–236. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19773-9_23
36. Luckin, R., Clark, W.: More than a game: the participatory design of contextualised technology-rich learning experiences with the ecology of resources. J. e-Learn. Knowl. Soc. **7**(3), 33–50 (2011)
37. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: a meta-analysis. J. Educ. Psychol. **106**(4), 901 (2014)
38. Miller, J.K., Friedman, B., Jancke, G., Gill, B.: Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. In: Proceedings of the 2007 International ACM Conference on Supporting Group Work, pp. 281–290. ACM (2007)
39. Mitchell, V., Ross, T., May, A., Sims, R., Parker, C.: Empirical investigation of the impact of using co-design methods when generating proposals for sustainable travel solutions. CoDesign **12**(4), 205–220 (2016)
40. Muller, M.J., Kuhn, S.: Participatory design. Commun. ACM **36**(6), 24–28 (1993)
41. Mulligan, D.K., King, J.: Bridging the gap between privacy and design. U. Pa. J. Const. L. **14**, 989 (2011)
42. Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A.K., Lee, M.K.: A fieldwork of the future with user enactments. In: Proceedings of the Designing Interactive Systems Conference, pp. 338–347. ACM (2012)
43. Olsen, J.: Orchestrating Combined Collaborative and Individual Learning in the Classroom. Ph.D. thesis, Carnegie Mellon University (2017)

44. Pane, J.F.: Informing progress: insights on personalized learning implementation and effects. RAND (2017)
45. Prieto, L.P., Holenko Dlab, M., Gutiérrez, I., Abdulwahed, M., Balid, W.: Orchestrating technology enhanced learning: a literature review and a conceptual framework. Int. J. Technol. Enhanced Learn. **3**(6), 583–598 (2011)
46. Prieto Santos, L.P., et al.: Supporting orchestration of blended CSCL scenarios in distributed learning environments (2012)
47. Ritter, S., Yudelson, M., Fancsali, S.E., Berman, S.R.: How mastery learning works at scale. In: Proceedings of the Third (2016) ACM Conference on Learning@ Scale, pp. 71–79. ACM (2016)
48. Ritter, S., Anderson, J.R., Koedinger, K.R., Corbett, A.: Cognitive tutor: applied research in mathematics education. Psychon. Bull. Rev. **14**(2), 249–255 (2007)
49. Ritter, S., Yudelson, M., Fancsali, S., Berman, S.R.: Towards integrating human and automated tutoring systems. In: EDM, pp. 626–627 (2016)
50. Rodriguez Triana, M.J., et al.: Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. Int. J. Technol. Enhanced Learn. **9**, 126–150 (2017)
51. Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R.: Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. Learn. Instr. **21**(2), 267–280 (2011)
52. Schofield, J.W.: Psychology: computers and classroom social processes-a review of the literature. Soc. Sci. Comput. Rev. **15**(1), 27–39 (1997)
53. Schofield, J.W., Eurich-Fulcer, R., Britt, C.L.: Teachers, computer tutors, and teaching: the artificially intelligent tutor as an agent for classroom change. Am. Educ. Res. J. **31**(3), 579–607 (1994)
54. Sharples, M.: Shared orchestration within and beyond the classroom. Comput. Educ. **69**, 504–506 (2013)
55. Sheridan, T.B.: Function allocation: algorithm, alchemy or apostasy? Int. J. Hum.-Comput. Stud. **52**(2), 203–216 (2000)
56. Trischler, J., Pervan, S.J., Kelly, S.J., Scott, D.R.: The value of codesign: the effect of customer involvement in service design teams. J. Serv. Res. **21**(1), 75–100 (2018)
57. Walker, E., Rummel, N., Koedinger, K.R.: Adaptive intelligent support to improve peer tutoring in algebra. Int. J. Artif. Intell. Educ. **24**(1), 33–61 (2014)
58. Walonoski, J.A., Heffernan, N.T.: Prevention of off-task gaming behavior in intelligent tutoring systems. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 722–724. Springer, Heidelberg (2006). https://doi.org/10.1007/11774303_80
59. Watters, A.: The Monsters of Education Technology. Smashwords Edition (2014)
60. Wetzel, J., et al.: A preliminary evaluation of the usability of an AI-infused orchestration system. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10948, pp. 379–383. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_71
61. Williamson, B.: Calculating children in the dataveillance school: personal and learning analytics. In: Surveillance Futures, pp. 62–90. Routledge, London (2016)
62. Williamson, B.: Decoding classdojo: psycho-policy, social-emotional learning and persuasive educational technologies. Learn. Media Technol. **42**(4), 440–453 (2017)
63. Wong, R.Y., Mulligan, D.K., Van Wyk, E., Pierce, J., Chuang, J.: Eliciting values reflections by engaging privacy futures using design workbooks (2017)
64. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-aware tutors: recognising and responding to student affect. Int. J. Learn. Technol. **4**(3–4), 129–164 (2009)

65. Yacef, K.: Intelligent teaching assistant systems. In: 2002 Proceedings of International Conference on Computers in Education, pp. 136–140. IEEE (2002)
66. Zhu, H., Yu, B., Halfaker, A., Terveen, L.: Value-sensitive algorithm design: method, case study, and lessons. Proc. ACM Hum. Comput. Interact. **2**(CSCW), 194:1–194:23 (2018). https://doi.org/10.1145/3274463, http://doi.acm.org/10.1145/3274463
67. Zimmerman, J., Forlizzi, J.: Speed dating: providing a menu of possible futures. She Ji: J. Des. Econ. Innov. **3**(1), 30–50 (2017)