Computers and Education Open 7 (2024) 100203

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Accuracy and effectiveness of an orchestration tool on instructors' interventions and groups' collaboration

LuEttaMae Lawrence [a,*], Emma Mercier [b], Taylor Tucker Parks [b], Nigel Bosch [b], Luc Paquette [b]

[a] *Utah State University, USA*
[b] *University of Illinois Urbana-Champaign, USA*

ARTICLE INFO

Keywords:
Collaborative learning
Human-computer interface
Teaching/learning strategies

ABSTRACT

This paper presents the development of a novel orchestration tool that predicts collaborative problem-solving (CPS) behaviors of undergraduate engineering groups and investigates the use of that tool by instructors. We explore the impact of receiving real-time, machine-learning, model-based prompts on 1) instructors' orchestration strategies, which are strategies instructors use to manage and facilitate collaborative activities, and 2) groups' participation, including how groups are engaged in CPS activities. The orchestration tool is a dashboard that notifies instructors of—and advises them on—monitoring and intervening with groups who may need collaborative support and guidance. We describe the accuracy of the models in predicting CPS behaviors and of instructors in identifying these behaviors in the classroom. We then describe how real-time prompts from models can affect instructors' orchestration strategies and students' participation. Our findings show that there is variability in the accuracy of our machine learning models and that instructors are better at identifying predictive behaviors as compared to the models. Instructors in this context engaged in orchestration strategies, like monitoring and probing when using the orchestration tool, and groups of students were largely talking while on-task across classes. We triangulate across data sources to examine the effectiveness of the orchestration tool in the classroom and share pedagogical and technical implications for the field.

Introduction

Collaborative problem-solving (CPS) is a crucial 21st-century skill [1]. This form of learning refers to joint activities completed by groups of two or more people to solve a problem without an obvious or clear solution [2,3]. Social knowledge construction, as proposed by Vygotsky [4], emphasizes the role of social interaction in learning environments. While theorists argue that social interaction is critical to developing disciplinary knowledge, these interactions do not occur naturally in classrooms. Learning environments must be intentionally created to foster social interaction, knowledge creation, and problem-solving skills [4]. Teachers play a vital role in creating environments conducive to collaborative learning to create and clarify the processes of social interactions [5]. While teacher guidance is essential, identifying when and how to intervene is challenging, especially in CPS tasks where groups may be working on different parts of the task simultaneously [6].

One way to support teachers in creating a classroom context that promotes social construction of knowledge is to leverage actionable analytics to help teachers orchestrate CPS tasks. While researchers have developed and tested analytics-rich tools to facilitate peer interactions, these tools often focus on monitoring or understanding collaborative interactions and are rarely implemented in face-to-face classroom [7]. More specifically, recent reviews on supporting collaboration using analytics call for more support for teachers to facilitate group discussion, as most collaborative analytics still rely on mirroring interaction rather than designs that enable alerting and advising [8]. In this paper, we introduce a Design-Based Implementation Research (DBIR) project that developed an orchestration tool using machine learning to provide real-time support for instructors during CPS activities. We tested how accurately machine learning models were able to predict groups' participation as well as instructors' ability to identify predicted behaviors. We then examined instructors' orchestration strategies while using the tool and how these strategies affected groups' participation. The contributions of our paper include (1) the development and testing of a

The authors declare that they have no conflict of interest.This work was supported by the [anonymized for review] [grant number].

* Corresponding author.
E-mail addresses: lu.lawrence@usu.edu (L. Lawrence), mercier@illinois.edu (E. Mercier), tdtucke2@illinois.edu (T. Tucker Parks), pnb@illinois.edu (N. Bosch), lpaq@illinois.edu (L. Paquette).

https://doi.org/10.1016/j.caeo.2024.100203
Received 16 February 2024; Received in revised form 3 June 2024; Accepted 18 July 2024
Available online 21 July 2024

novel real-time machine learning model and prompts for supporting instructors' interventions, (2) evidence of potential interventions in machine-learning supported classrooms, and (3) a DBIR approach and detailed analytic process for triangulating data types to understand the tool's impact.

Teacher guidance in collaborative problem-solving (CPS)

Research has shown that CPS allows students to deepen their knowledge and develop better problem-solving skills [3,9]. CPS relies on social knowledge construction, where groups of students engage in synchronous activity to construct and maintain a shared understanding of a problem [3,4]. To engage in social knowledge construction, students must disseminate their own knowledge while negotiating their group's shared understanding [1]. Researchers emphasize the need for groups to establish common ground, develop a solution, monitor their progress, and negotiate ideas [9] while working within a pre-constructed joint problem space [3]. Yet, to foster productive CPS, the learning environment must be designed to support and augment these interactions, including the teachers' facilitation [10].

Teachers play a crucial role in planning and facilitating CPS [5,11]. Kaendler and colleagues [5] theorize that to foster high-quality interactions, teachers must plan, monitor, support, and reflect on groups' interactions. Specifically, teachers need to identify the nature of a group's interactions to intervene when necessary to avoid interrupting productive interactions [6], as teachers' interventions has been shown to influence the quality of group interactions [12]. Webb and colleagues [12] identified ways teachers guided students to collaborate in interventions, including asking for explanations, sharing partial answers or directions, requesting elaborations, and prompting students to discuss content. While these strategies for orchestrating activities are important for facilitating productive collaboration, it is a challenge to identify when a group may need support and what type of guidance is needed [6].

Research indicates that without training, orchestration strategies, such as monitoring, probing, and prompting collaboration, do not occur naturally [13], even expert teachers feel unprepared for such activities (e.g. [11]). In our context, teaching assistants (TAs), who frequently teach core undergraduate courses, are rarely equipped with the pedagogical strategies needed to facilitate effective CPS [6]. Research shows that graduate and undergraduate TAs often lack the pedagogical knowledge to monitor, assess, and support groups' interactions in real-time [6,14]. Our previous research in engineering courses shows that TAs rarely intervene in ways that promote CPS, and instead disrupt moments of productive collaboration [14]. Because CPS provides opportunities to enhance disciplinary knowledge, social interactions, and problem-solving skills, it has become an increasingly common pedagogical practice in postsecondary courses and instructors need support to enact it well (e.g., [15]). Given the challenges TAs have monitoring and intervening in ways that support CPS [6,14], we leverage an orchestration tool to provide instructors with real time support to learn about and facilitate productive interactions.

Analytics-Rich orchestration technology

Orchestration, defined as coordinating interventions across learning activities, is crucial in managing groups working simultaneously on different tasks [16]. Orchestration is an instructional method that often includes technological enhancements that call attention to specific moments in the classroom [17,18]. This form of targeted scaffolding often involves focusing teachers' attention on students or groups who may need assistance [19]. In doing so, real-time orchestration technologies support teachers at the interactive phase of Kaendeler and colleagues' [5] model, which includes monitoring and pedagogical decision-making.

There is a growing consensus that keeping teachers up to date on groups' collaborative progress can lessen teachers' load and afford them more time to focus on providing support rather than trying to identify who may benefit from support [20]. To do so, many orchestration tools leverage log-based analytics to explain or characterize groups' interactions that are then fed back to teachers (e.g., [21]). However, visualizing and sharing log data does not guarantee that teachers will be able to interpret and act on it [20]. Scholars argue that to develop actionable collaborative analytics, they must be theoretically grounded, co-created with users, and built from well-designed tasks that capture meaningful data [22]. Processes to co-design analytics based on actual needs and learning goals are hypothesized to lead to more authentic and actionable insights [22,23].

Recent reviews have called to move beyond descriptive analytics and visualizations and toward more actionable recommendations that advise teachers regarding who to help and why [7,8]. However, actionable recommendations should not restrict pedagogical goals. To provide teachers with guidance to foster productive CPS, analytics should offer advice for interaction rather than directive instruction. For instance, analytics that capture what is happening, in conjunction with scaffolding (such as sentence openers or probing questions) can provide teachers with actionable prompts that support their interventions [24]. However, teachers need to retain their autonomy in being able to make pedagogical decisions based on information from analytics [25]. Thus, critical design concerns of orchestration technologies include providing teachers with guidance and potential actions while also reinforcing their agency to question the analytics and make decisions based on what they observe in the classroom [26]. While this is a growing body of literature in the CSCL field, few studies have yet to examine how orchestration tools might support teacher guidance toward groups' interactions [8]. Our work builds on this existing work to examine how an orchestration tool, developed with instructors, can support their interactions and, in turn, groups' participation in CPS activities.

Research questions

This study addresses open questions about the accuracy, effectiveness, and use of model-based prompts in orchestration tools during CPS activities. Specifically, we look at how the CSTEPS tool, including features that prompt from machine learning models and advise instructors on how to monitor and intervene, impact groups' participation We hypothesize that predictive models embedded in orchestration technology, when utilized by teachers, can improve instructors orchestration strategies (i.e., monitoring, intervening, attention; see Fig. 1) which can in turn engage students in participating in the CPS activity. To investigate this hypothesis, we address each of the following questions:

1) How accurate were the models at predicting groups' participation?
2) What was the relationship between the model predictions, the expert coder's analysis of behaviors, and the instructor's identification of behaviors?
3) How did the *CSTEPS tool* support instructors' orchestration strategies in the classroom?
4) How did instructors' use of the *CSTEPS tool* and orchestration strategies support groups' participation?

Methods

Study design and background

This study is part of a multi-year, DBIR project [27] called *CSTEPS* (Collaborative Support Tools for Engineering Problem Solving). We built upon several iterations of our DBIR process, in which we explored the learning context, co-created tasks to foster collaboration [28], trained TAs on CPS [29], built and evaluated student collaborative tools [30], co-designed the orchestration tool with TAs [31], and developed models to predict CPS behaviors [32].

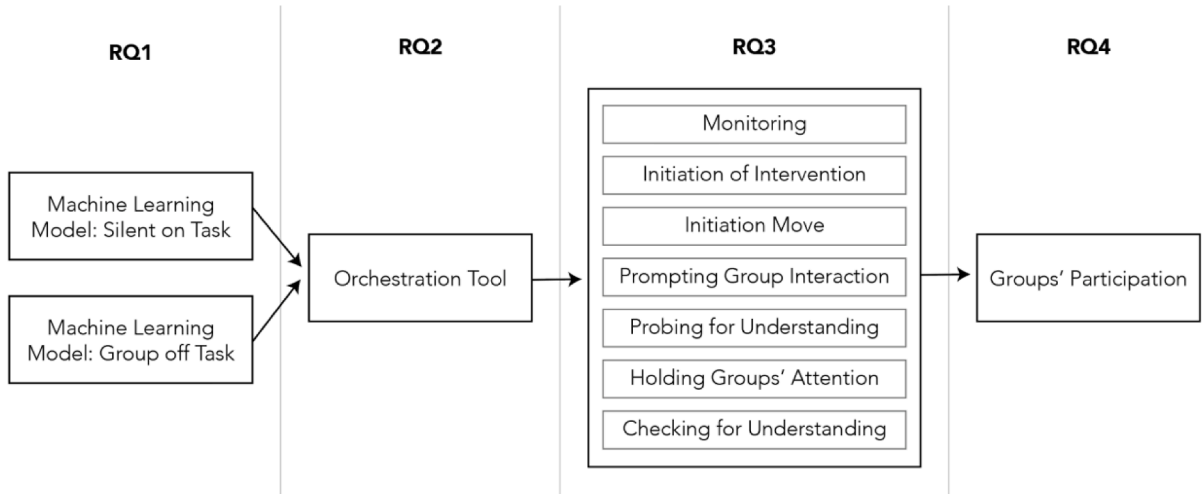


Fig. 1. The conceptual model outlining our research questions and their interrelationships.

Context and participants

Data were collected from 90 undergraduate engineering students (20 female, 70 male) enrolled in an introductory engineering course during the Spring 2019 semester in a lab-based classroom. The students were organized into 26 groups across five discussion sections. Groups were formed using Comprehensive Assessment of Team Member Effectiveness (CATME) to ensure no students were a minority in the group (e.g., not one female with three males). Each student in the group used synchronized tablet-based applications for collaborative tasks. These tasks were co-designed by members of the research team using a literature-based framework to support CPS [28] (see Appendix A for example). Each one presented students with ill-structured, real-world applications of key concepts to resemble a workplace problem, with multiple solutions that were socially negotiable through multiple solution paths [3,28] . Each group produced one worksheet with their joint solution. The study involved five discussion sections, each led by a graduate teaching assistant (TA) and two undergraduate course assistants (CAs). The two TAs and six CAs (Table 1) are henceforth called instructors. Consent was obtained from all participants. The course took place in a lab-based, university classroom, where all classes were held for the full duration of the semester.

Students' technology & predictive models

Students used a tablet-based application that facilitated the joint task by displaying each other's work to the group (Fig. 2). Students were trained on the features of the software in week one, and used it through the remainder of the semester, making it a normal part of their learning experience. This synchronized software allowed us to collect log data about students' interactions during the task. This data, in conjunction with the coding of the groups' participation, was used to inform the machine learning models used in the orchestration tool. Our CPS machine learning models were created from video and log file analysis of 20 groups across six weeks in a previous phase of the project [32]. Full details of our model-building process are available in the results from the previous phase of the project [32].

Table 1
Participants by section.

	Section 1	Section 2	Section 3	Section 4	Section 5
Number of Groups Analyzed	5	6	3	6	4
Number of Students Analyzed	18	22	10	22	14
TA	Adam	Adam	Adam	Lisa	Lisa
CA 1	Zain	Jason	Jason	Casey	Raj
CA 2	Santu	Jenn	Jenn	Santu	Santu

CSTEPS tool

The *CSTEPS tool* was co-designed over two iterations [31]) and implemented in Javascript and HTML (available online at https://github.com/colearnlab/ccaf-web). Machine learning models were trained to alert instructors when a group was likely to need support and provide intervention strategies (Authors). Specifically, we fit random forest classification models with nested cross-validation for hyperparameter tuning using *scikit-learn* in Python [33]. Trained models and code for running models in a server application to make real-time predictions are available (https://github.com/colearnlab/ccaf-prediction-server). The *CSTEPS tool* presented all groups in a card format with a color bar along the top (see Fig. 4). Yellow indicated that there was no prompt, meaning that the prediction models had not identified a potential issue among the group. The orange bar meant there was a prompt. When another instructor in the classroom selected a prompt, the bar turned gray and was not selectable. Each card showed a thumbnail of the current worksheet and the students' location. Instructors could view a group's work, allowing them to join the worksheet to monitor and write in the shared drawing space.

When an instructor opened a prompt, a pop-up prompt would appear on the screen (see Fig. 5). On the left side, the tool prompted the instructor to monitor the group for the predicted behavior. When prompted to monitor a group, the tool provided the instructor with a title, a drop-down menu for more information, and buttons to either confirm or deny if the models had accurately described the group's behavior (Fig. 3). If the instructor declined the prompt (i.e., if they saw that the model did not accurately reflect the group's behavior and rejected its suggestion), the pop-up disappeared. If they confirmed the issue, the tool then used the right panel to present strategies for intervention. Strategies were co-designed and organized into two sections: 1) an overarching goal that identified a strategy for approaching the group, and 2) a few sentence starters or phrases to use when talking to the students. Once their intervention was complete, the instructor could exit the prompt by selecting the "Done" button.

Data collection

Data Collection: This Design-Based Implementation Research (DBIR) study employed a mixed-methods approach, combining log data,

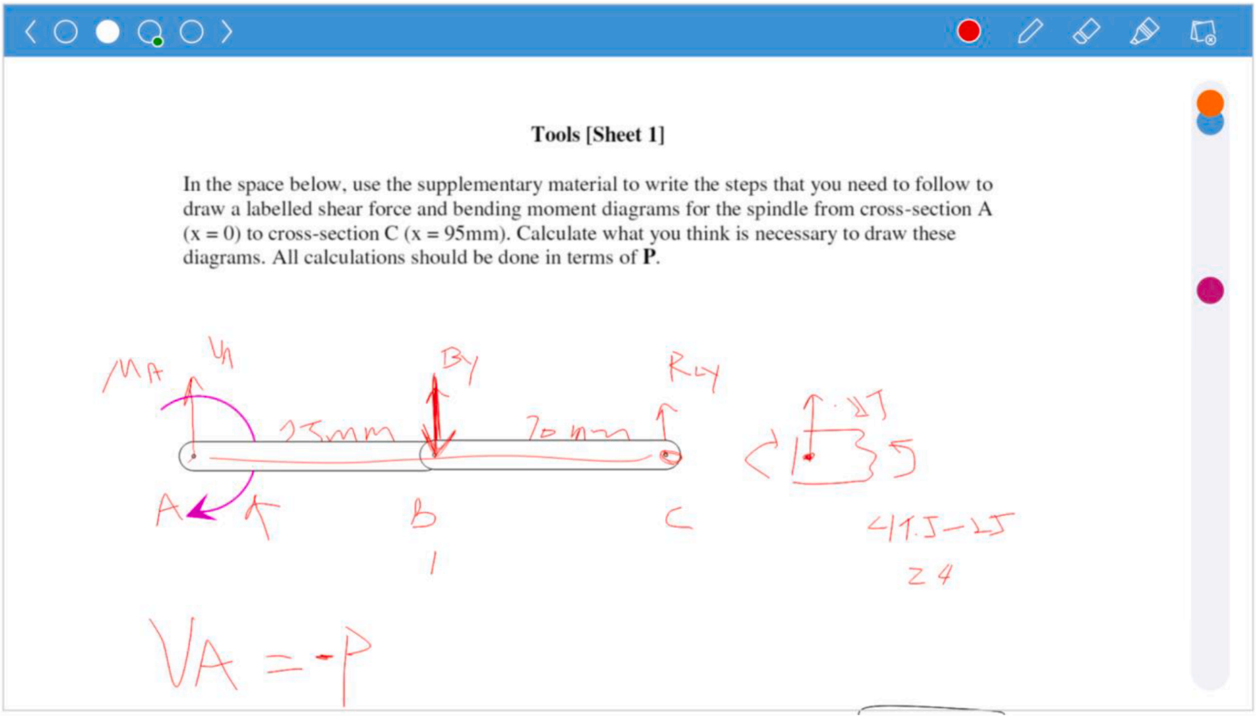


Fig. 2. The student tool included navigation of pages in the worksheet and location of students on the pages, a series of drawing tools, the PDF worksheet pages where the students were able to draw together, and a scrollbar indicated the location of students on the page.

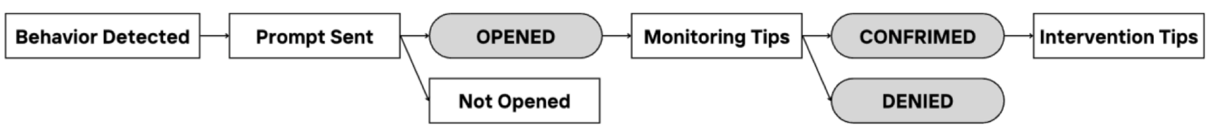


Fig. 3. Flow chart of instructors' use of the CSTEPS tool.

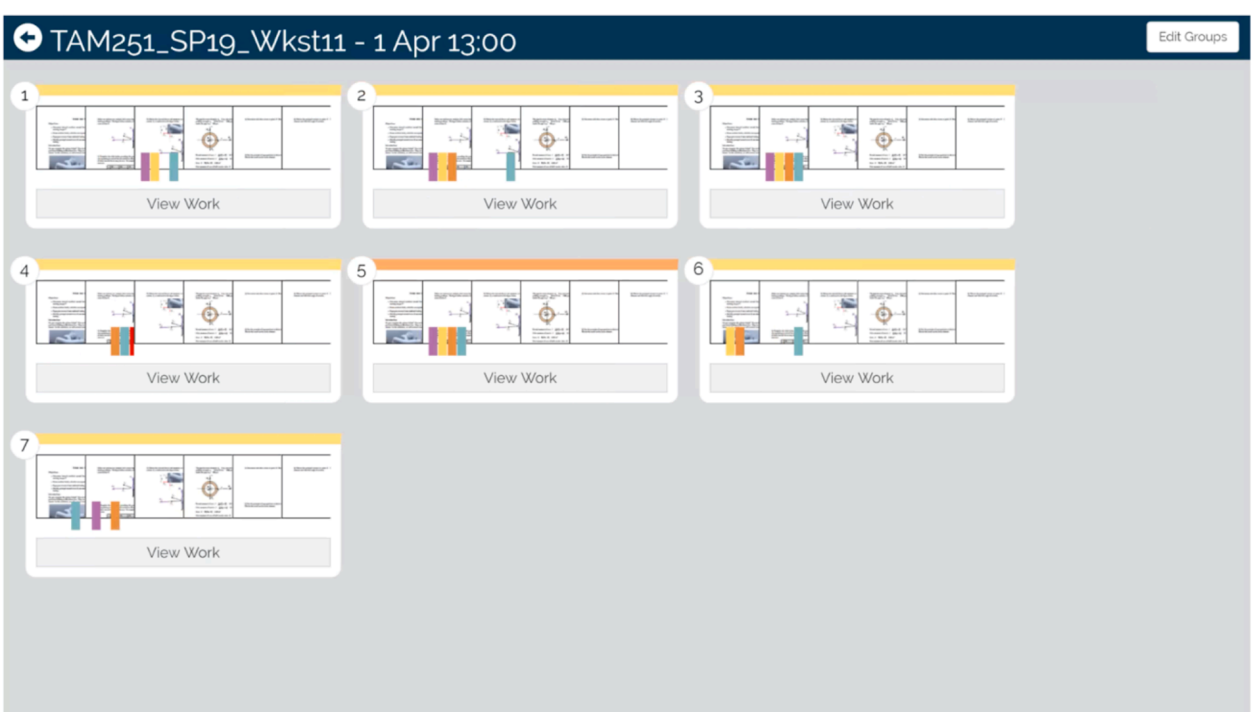


Fig. 4. *CSTEPS tool*, main screen view.

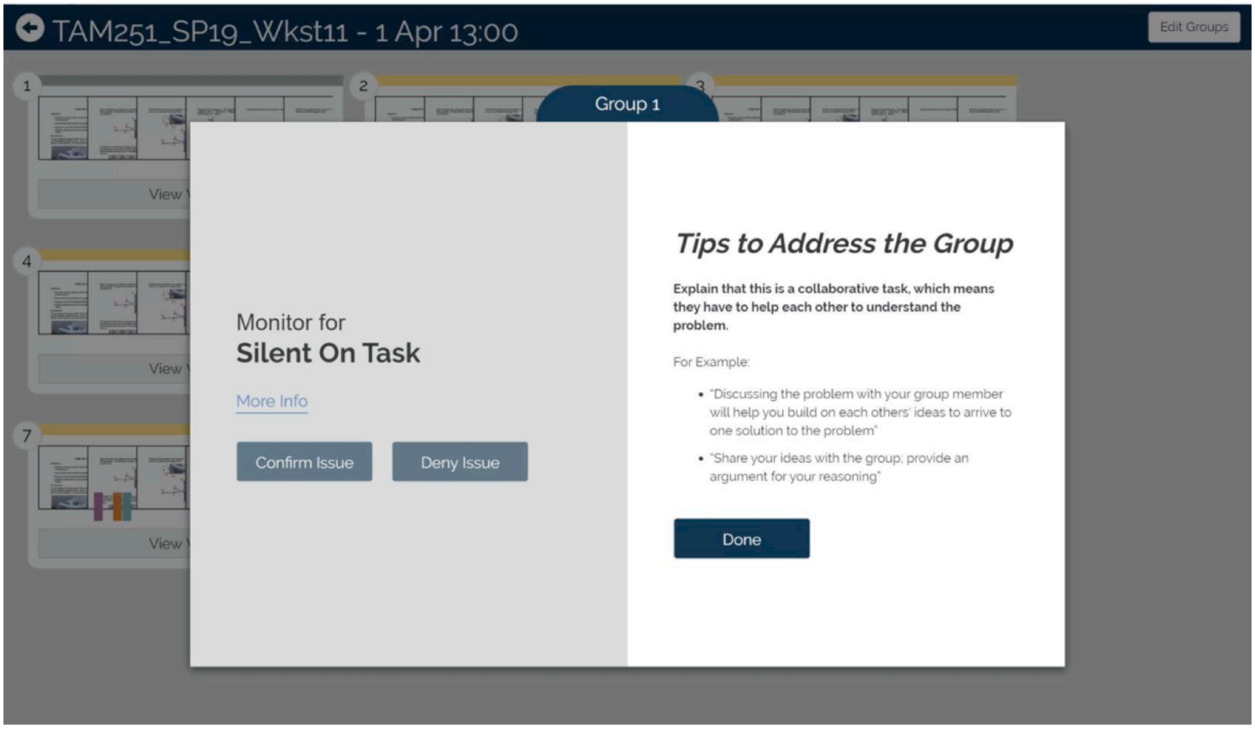


Fig. 5. *CSTEPS tool* after an instructor has selected a group. On the left, the tool prompts the instructor to monitor the behavior. If the prompt is confirmed, the right panel displays tips to address the group.

video recordings, and audio recordings to triangulate data and gain a more complete picture of the classroom [34]. Data were collected during a 3-week implementation, weeks 10, 11, and 12 of the 16-week semester. Instructors were trained on the tool functions prior to implementation, and students were introduced to the tool the first week it was used. Log data included timestamps for prompt visibility, selection, and confirmation/denial by instructors; all log data included timestamps and instructor names. Audio and video data were collected from all consenting participants in the lab classroom; students that did not give consent were grouped together and were not video or audio recorded. Overhead cameras and either a hanging or wireless microphone collected data from each table in the classroom and instructors wore lapel microphones. We received approval from our Institutional Review Board for data collection, for which we followed protocols about data storage (i.e., storing data in a secure Box folder), participant privacy (i.e., anonymizing data), and other ethical issues.

Analysis

First, we broke up the video data into 20-second clips, the same unit of analysis the machine learning models used to predict group behaviors [32]. We chose the 20-second length after rigorously testing video coding with different clip lengths (i.e., 5, 10, 20, 30, & 60 s). We determined that 20 s was the best unit of analysis to reliably identify what the group was doing without observing conflicting behaviors within the same clip. Coding rubrics included identifying whether the group was silent or speaking and what type of dialogue was occurring, and if their behavior (and dialogue when applicable) was on-task or off-task. It is important to note that the synced tablets represented a medium through which students could silently collaborate. Thus, groups could be *silent on task, talking on task, silent off-task*, or *talking off-task* (*IRR* = 0.65; 88 % agreement). Next, we aligned the video and log file data segments to illuminate the relationships between data. We examined the three 20-second clips prior to (1) the models showing a prompt in the tool and (2) an instructor responding to a prompt (e.g., confirming or denying it).

To analyze the instructors' interactions with groups, interventions were identified in the video data and transcribed [35]. Interventions were defined as moments when an instructor spoke with a group. Each intervention was coded for the presence of orchestration strategies (see Appendix B for definitions and reliability). Each transcribed intervention was framed by three 20-second clips of student dialogue before and after to enable us to observe the change in the group's behavior. This matched the *Binded* tool which also used three-clips to identify group behaviors (i.e., a behavior must be identified for three consecutive 20-second clips, or 60 s, to generate a prompt in the tool). We used several analytic approaches to explore our four research questions:

Research Question 1: To determine the accuracy of the models, we analyzed the retroactively coded video clips against the predictions generated from log file data. This comparison allowed us to identify whether the behaviors that the machine learning models prompted matched the participation behaviors coded by our research team. The alignment was coded as accurate if at least one of the three 20-second video clips coded by the research team matched the prompt in the tool. We calculated descriptive statistics to measure the alignment of these data.

Research Question 2: We investigated the relationship between the model predictions, the expert coder's analysis of behaviors, and the instructor's identification of behaviors. We computed descriptive statistics of the alignment between the instructor's identification—selecting "confirm" or "deny" to a prompt shown in the orchestration tool—to the coding of the group's behavior at the time the instructor answered the prompt. We then examined the relationship between the accuracy of the models and the instructors.

Research Question 3: To understand how the orchestration tool affected instructors' orchestration strategies during interventions, we compared the occurrence of orchestration strategies with and without the orchestration tool. To be classified as having used the tool, the instructor had to interact with the tool for 60 s before or during an intervention; these were identified using timestamps in the log file data and validated with video data. We then ran *t*-tests between the instructors' use of orchestration strategies with and without the orchestration tool to identify any differences. Strategies and tool use were coded as dichotomous variables, which satisfy test assumptions by yielding approximately normal sampling distributions when $n \times p \geq 5$ and $n \times (1 - p) \geq 5$. Given few closely related statistical tests, we did not control the false discovery rate for multiple tests (additionally, doing so would not substantially affect interpretation as all but one significant results were highly significant).

Research Question 4: Finally, we examined the clip coding of a group's behavior before and after interventions to examine how the instructor's use of the *CSTEPS tool* and implementation of orchestration strategies supported the group's participation. Our expert clip coding was categorized by *silent on task, talking on task, silent off-task,* and *talking off-task* to understand how the behaviors in the tool were actualized in the classroom. We cluster these into two categories: *talking while on task* and *silent or off-task* (*silent on task, silent off-task,* and *talking off-task*) to look at the transitions between categories before and after instructor interventions. We examined groups' overall transitions between the behaviors predicted in the orchestration tool before and after interventions, as well as any interactions between tool use and the presence of orchestration strategies.

Results

RQ 1. How accurate were the models at predicting groups' participation?

We conducted an initial pilot study prior to giving the *CSTEPS tool* to instructors. We monitored a test version of the teacher tool while students engaged in their regular classroom activities to verify that the prompt infrastructure worked. However, no prompts were triggered during the pilot. This may have been the result of the initial models' predictions having been calibrated to the training data, which were collected from different students in another semester. The older data may have caused the models to have lower confidence in their predictions in this new context.

To counteract the models' low confidence values, we adjusted the thresholds required for the detection of off-task and silent on-task behavior. We systematically changed the thresholds to identify how many prompts would have been triggered during a class under the new thresholds. Following this analysis, we adjusted the threshold for off-task behavior to 0.15 and the threshold for silent on-task behavior to 0.40. By lowering these thresholds, we created a situation in which the models were more likely to generate false positive predictions, leading to a higher number of prompts generated during each class. We accepted this overprediction of the target behaviors for two key reasons. First, one of the tool's goals was to train instructors to recognize off-task and silent on-task behavior, and it always asked instructors to first verify that the behavior was currently ongoing before offering guidance. Second, given that the prompts were targeting rare behaviors, we preferred overprediction of the target behaviors to underprediction, which would have missed most of the moments where an instructor's intervention might have been helpful.

Across all five classes, the machine learning models generated 437 prompts in the orchestration tool. The clip coding alignment indicates that 134 (30 %) of the models' prompts were confirmed by our video coding. Of the total 437 prompts, 213 alerts predicted that a group was off-task and 224 alerts predicted that a group was silent on task. 34 of the 213 (14 %) of the off-task prompts aligned with the video coding, whereas 100 of the total 224 (45 %) silent on-task prompts were confirmed. This suggests that the models' predictions of silent on-task behavior more frequently matched our video coding. In general, the models over-predicted the occurrence of behaviors requiring monitoring, which was expected and intended to avoid missing key events that should be predicted (i.e., false negatives).

RQ 2. What was the relationship between the model predictions, the expert coder's analysis of behaviors, and the instructor's identification of behaviors?

Of the 437 generated prompts in the orchestration tool, instructors opened 232. Because this is a supportive tool and not intended to be prescriptive, we expected prompts to go unopened. Instructors moved from group to group for many reasons, including prompts from the orchestration tool, observations, and student questions. They did not have time to open all prompts. There were differences regarding how instructors used the technology. Adam, Santu, Jenn, and Lisa opened the most prompts, whereas the remaining teachers opened far fewer (Fig. 6).

Of the 232 prompts the instructors opened, 161 (70 %) matched our coding of the video clips; more specifically, 103 of the 232 prompts were group off-task prompts, 76 (74 %) of which were identified correctly by the instructor, and 129 were silent on-task prompts, 85 (67 %) of which were identified correctly. After opening the prompts, instructors confirmed or denied that the behavior was present in the group. Of the total 232 opened prompts, 59 (25 %) were confirmed and 173 (75 %) were denied. Of the 59 confirmed prompts, 23 (39 %) were correctly identified. Five (19 %) group off-task prompts were correctly identified, and 18 (55 %) silent on-task were correctly identified. This shows that instructors were slightly better at identifying the presence of silent on-task behaviors correctly compared to group off-task behaviors. In contrast, of the 173 prompts that were denied by instructors, 138 (80 %) were correctly identified by instructors. Seventy-one (92 %) were correctly identified as group off-task behaviors and 67 (70 %) as silent on-task behaviors. This data illustrates that the instructors were better at correctly indicating prompts when they were denying behaviors and were more likely to incorrectly agree with the prompts. Instructors' proportion of correctly identified behaviors ranged from 60 % to 80 % (M=69 %, SD = 7 %; Table 3). There was no relationship between prompts opened and correctness, as the two instructors with the highest level of accuracy were the instructors who opened the fewest and most prompts respectively.

To compare the accuracy of the machine learning models, the instructors, and the video clip coding, we examined the relationship between these aligned data. Of the 232 prompts that were confirmed or denied by an instructor, 73 of them matched the log data's level of accuracy and 159 did not match. This means that the model, the instructor, and the expert coders agreed on 31 % of the prompts that were opened and answered. Of the 73 prompts for which the model, instructor, and coder identified the same behavior, 47 were silent on-task and 26 were group off-task.

While this is a low frequency of matching between all parties, we acknowledge that very few prompts were responded to immediately after being shown. On average, prompts were answered 55 s after they were opened (SD = 70 s; $Mode$ = 60 s), with a range from 5 s to 9 min. If a prompt was answered 60 s later, the instructors observed different segments of time, and our analysis examined different clip codes. Therefore, the model could identify something incorrectly, whereas the instructor identified it correctly. For example, the model might send a prompt for silent on-task behavior, yet our clip coding could disagree. If 60 s later, the instructor confirms silent on-task behavior and the clip coding confirms that they were correct, that would indicate that the group's behavior changed between the time the prompt was generated and when the instructor monitored for that behavior.

RQ 3. How did the CSTEPS tool support instructors' orchestration strategies in the classroom?

The prompts in the tool were the focus of the tool, and the secondary use of the technology was to join a group and view their work. This function was used in a variety of ways unrelated to the number of prompts an instructor opened (see Fig. 6). To understand how the *CSTEPS tool* supported orchestration strategies, we analyzed all

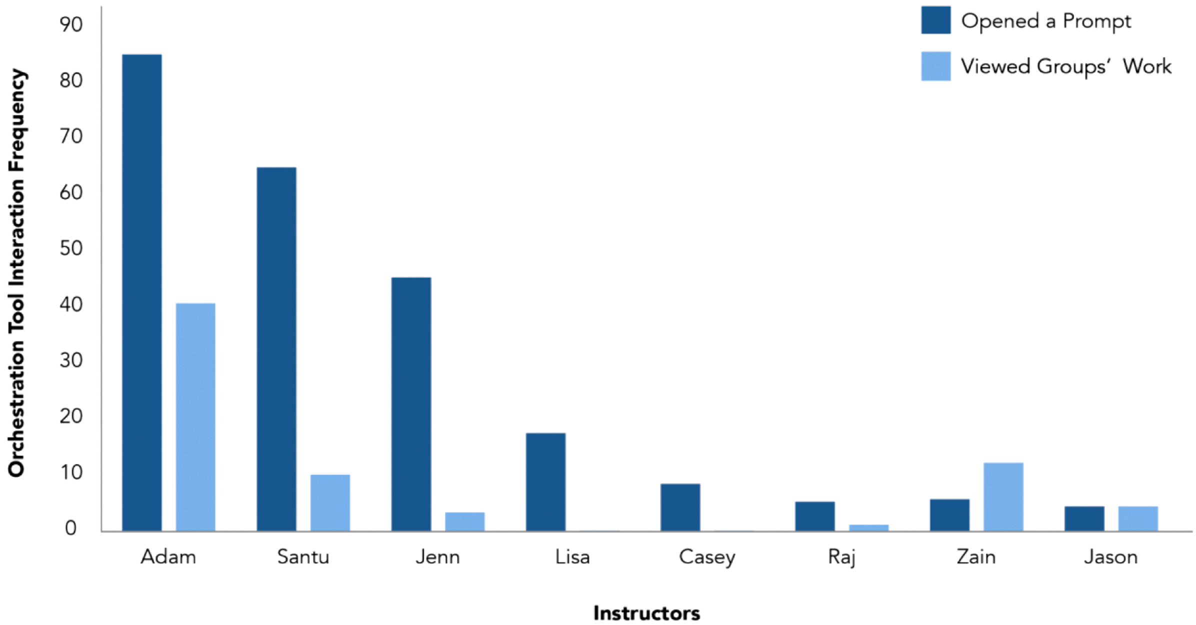


Fig. 6. Orchestration tool interaction frequency separated by instructors with both prompts opened and view groups' work.

Table 3
Instructors' accuracy of prompts.

Instructor	Prompts Opened	Prompts Confirmed	% Prompts Confirmed	Prompts Identified Correctly	% Identified Correctly
Adam	84	22	26 %	64	76 %
Lisa	17	4	23 %	11	65 %
Casey	8	0	0 %	5	63 %
Jason	4	0	0 %	3	75 %
Jenn	45	19	42 %	29	64 %
Raj	5	0	0 %	4	80 %
Santu	64	11	17 %	42	65 %
Zain	5	3	60 %	3	60 %
Total	*232*	*59*		*161*	

interventions, including both opening and responding to prompts as well as viewing a group's work.

Instructors intervened with groups 297 times (Table 4), with a range of zero to 40 interventions per section (M=20, SD = 11). Instructors used the *CSTEPS tool* before or during 57 interventions (19 %), including 50 prompts, five uses of the view work function, and two instances that included both a prompt and the use of view work. During validation with video data, we found that another instructor who was not interacting with the group used the technology in 21 of the 50 instances in which prompts were opened. All these prompts were denied by the

Table 4
Tool use during interventions.

Instructor	Interventions	Instructor Used Tool During ANY Intervention	Used Tool During Intervention THEY ENACTED	Tool Use During Someone else's Intervention
Adam	63	12	9	3
Lisa	61	8	1	7
Casey	11	3	0	3
Jason	52	6	4	2
Jenn	36	13	13	0
Raj	16	1	1	0
Santu	44	11	6	5
Zain	14	3	2	1
Total	297	*57*	*36*	*21*

instructors who opened them; they were counted as 'interventions without tool use' in analyses because they did not affect the intervening instructors' interactions.

Of the 36 instances in which the intervening instructor used the tool, five used the view work function, 14 responded to group off-task prompts, 15 responded to silent on-task prompts, and two responded to silent on-task prompts followed immediately by the use of the view work function. We also examined the timing of all tool use. In sixteen instances, the tool was used before the intervention, meaning the instructor opened and responded to the prompt or viewed a group's work before ever interacting with the group. Instructors also used the technology 10 times during their intervention; in 4 instances, instructors used the view work function and in 6 instances the instructor opened and responded to a prompt while still with a group. Finally, in 10 cases, an instructor opened a prompt before intervening and responding either during (n=2) or immediately afterward (n=8).

To understand how instructors interacted with groups, each intervention was coded for the presence of orchestration strategies. Fig. 7 illustrates the proportion of each strategy across interventions. Instructors were identified as "monitoring" when they spent at least 10 s observing a group of students before intervening. There was no significant difference for the use of monitoring as a strategy with (M=22 %, SD = 42 %) or without the orchestration tool (M=23 %, SD = 42 %), d = −0.02, t(60.9) = −0.12, p = .91. Interventions were either initiated by the instructor or the students. When using the tool, instructors initiated significantly more interventions (M=71 %, SD = 46 %) compared to without the tool (M=44 %, SD = 05 %), d=0.559, t(63.9) = 3.65, p < .01. Since the technology is designed to direct instructors towards certain groups, we hypothesize that instructors were more likely to initiate strategies during interventions without the technology. Our findings show that roughly half of all interventions began with the instructor using a probing initiation strategy rather than immediately providing direct instruction (M=48 %, SD = 50 %). There was a significant difference in initiation strategies between interventions in which instructors used the tool (M=66 %, SD = 48 %), which prompted them to use generative initiation moves, compared to those in which they did not use the tool (M=42 %, SD = 49 %), d=0.509, t(62.3) = 2.96, p < .01.

Instructors probed groups' understanding in the initiation move by asking what the group was working on, how they were doing, or how

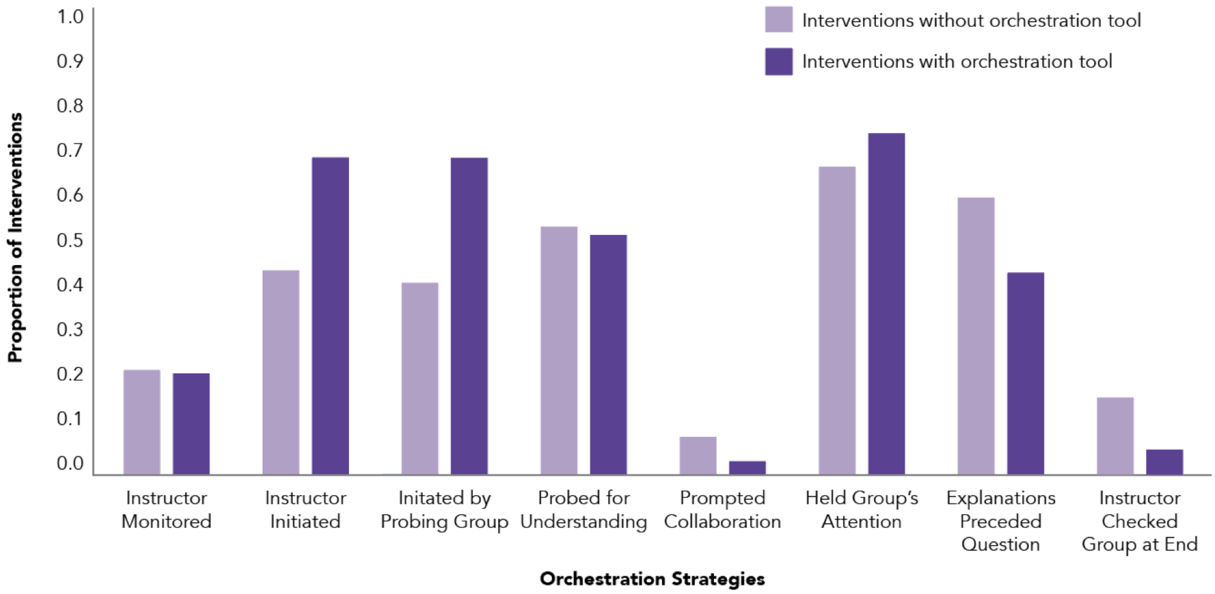


Fig. 7. Proportion of interventions organized by orchestration strategies with and without the orchestration tool.

much progress the group had made. In one example, Adam opened a group off-task prompt, walked over to the table, and monitored the group for a few seconds before confirming the prompt. The group quieted as Adam approached the table and asked, "*Having trouble with the worksheet?*" One member of the group, who was off-task, responded, "*I'm making… we're almost ready…*", followed by a few seconds of silence across the group. Another member of the group, who had not been engaged in off-task discussion said to Adam, "*So I have got a question…*" The student asked a question, which was followed by a clarification from Adam.

The interventions were additionally coded for follow-up moves to identify if instructors posed questions or provided direct instruction after the initiation move. Fifty-four percent (SD = 50 %) of interventions had follow-up moves in which the instructor further probed the understanding of the group, a non-significant difference with (M=50 %, SD = 49 %) and without the tool (M=55 %, SD = 50 %), $d = -0.13$, $t(60.2) = -0.76$, $p = .446$. Follow-up moves were also coded to identify moments wherein instructors asked questions or engaged the group in conversation, rather than telling them what to do or revealing the answer. In one case, Adam received a group off-task prompt about a group of three students. He opened the prompt, walked to the table, and monitored for six seconds. While the group was on task, Adam caught part of a conversation among the three students. Adam arrived at the table as Student 1 was asking a question to his group; this question was then posed to Adam as he moved toward the group.

Student 1 *Wait, for this… what could be the relation between* [Adam arrives] *like the X and the Y moments and shear stress are related… I don't know how they are related.*

Adam *What do we call a moment acting perpendicular to the direction?*

Student 2 *Torsion.*

Student 3 *It's like a twist.*

Adam *Torsion. Do we have a way to relate torque to a shear stress? You may not remember it was earlier in the semester…*

Student 1 *I don't, I don't…*

Adam *Look on the equation sheet, see if that will spark your memory.*

Rather than explaining how the moments and shear stress are related, or explicitly stating the group's next step, Adam provided guidance and helped the group identify how they might address the problem.

Next, we examined moments in which instructors directly prompted groups to work together. This strategy, wherein an instructor directly prompted the group to discuss or collaborate, rarely occurred in the data. There was no significant difference with the tool (M=7 %, SD = 25 %) compared to without it (M= 8 %, SD = 26 %), $d = -0.03$, $t(62.5) = -0.21$, $p = .833$. Despite this low frequency (n=22), all the instructors except Zain engaged in this strategy at least once. In some instances, the instructor repeated prompts that were supplied in the tool. In one instance, when Santu was answering a question, rather than providing a direct answer, he asked, "*Does anyone in your group know?*" In other cases, the instructors adapted strategies for their own use. For example, Adam prompted a group to work together by sharing advice: "*Maybe, I don't know, talk to each other. Work through it together. Even if only one of you is writing, you should be on the same problem.*"

Holding the group's attention aligns with the goal of promoting collaboration, wherein instructors can model engagement with multiple group members. Instructors held the full group's attention during the majority of all interventions (M=69 %; SD = 47 %), but no significant difference was observed between interventions with the tool (M=71 %, SD = 46 %) compared to without it (M=68 %, SD = 46 %), d=0.06, $t(61.4) = 0.38$, $p = .70$. Our coding identified whether explanations were preceded by questions, as opposed to the instructor explaining a concept without confirmation of what the group needed. Fifty-nine percent (SD = 49 %) of all interventions had at least one explanation prompted by a question or an expression of confusion from the group, with a barely significant difference with the tool (M=44 %, SD = 50 %) compared to without the tool (M=62 %, SD = 49 %), $d = -0.35$, $t(59.7) = -2.11$, $p = .04$. Finally, we coded instances in which instructors ended interventions by checking the group's understanding. There was little difference in the use of this strategy; instructors ended by checking for understanding with the tool in 9 % of cases (SD = 29 %) versus in 17 % of cases (SD = 37 %) without the tool ($d = -0.21$, $t(73.3) = -1.59$, $p = .12$). Closing knowledge checks were often simple questions posed to the group, such as "*Does that answer your question?*" or "*Did that make sense?*"

RQ 4. How did the TAs' use of the CSTEPS tool and orchestration strategies support groups' participation?

We analyzed the clip coding alignment before and after each intervention to understand whether the instructors' intervention and tool use affected students' participation. Looking at all 297 interventions, 94 % (n=278) were talking and on task before the intervention started. After an intervention occurred, 90 % (n=267) were talking and on task. Of the groups that were talking and on the task before the instructor intervened, the majority sustained this behavior afterward (n=255, 86 %), while a few transitioned to silently working or off-task behavior (n=12, 4 %, see Fig. 8). Seventy-seven percent of groups that were silent or off-task pre-intervention transitioned to talking while on task post-intervention (n=23). Our analysis showed that groups more frequently remained talking and on task when the orchestration tool was not used (87 %) compared to when it was used (78 %). The tool was used more often in cases where, after the instructor intervened, the group transitioned either from silent or off-task to talking and on task or vice versa.

Due to the highly unequal distribution across categories, we did not statistically analyze the orchestration strategies and tool use during the interventions in relation to group participation. Fig. 9 shows the proportions of orchestration strategies by tool use. In examining frequencies, we gleaned several high-level insights. Instructors monitored the groups more frequently during interventions wherein the group moved from a silent or off-task to talking and on task (50 %). Although such cases were rare, the instructors who used the tool monitored more frequently (100 %) compared to those without (40 %). Our findings show that, when the instructor initiated the intervention, the group was most likely to move from talking and on task to silent or off-task; they also used more direct instruction during the initiation move when the group remained silent or off-task. Finally, the instructors probed for understanding most frequently when the group moved from a silent off-task to talking and on task.

Discussion

In this paper, we present an exploratory study to leverage machine learning techniques to support real-time predictions of collaboration interactions to inform instructors' pedagogical moves. We investigated the accuracy and effectiveness of our novel orchestration tool by examining the relationship between machine learning models predicting CPS, instructors' responses to these predictions, instructors' orchestration strategies, and the effect on group participation. By looking at the interplay of these factors in an authentic classroom context, we can understand how orchestration technology could support instructors and groups in real-time. Below we describe our findings by contribution, including technological and pedagogical findings and implications.

Technological findings and implications

When analyzing the log file data alone, our findings show that the predictive models often identified participation inaccurately but were better at identifying *silent on-task* behaviors than *off-task* ones. We attribute this to several factors. First, we trained models on data from the same course but with different students in a previous semester. Indeed, this study aimed to understand how these models scale to other students. Second, we intentionally over-predicted the target behaviors to avoid missing crucial intervention opportunities for instructors to monitor and identify behaviors. Given the goal of training the instructors to practice monitoring, we considered inaccuracies acceptable and designed the interaction to ask instructors to verify predictions. Our tool's design enabled instructors to question the analytics and verify behaviors before conducting interventions that may not have been necessary [31].

When we looked at the instructors' use of the orchestration tool, we found differences in how they used the technology. We acknowledge that the one instructor who engaged with the tool the most was Adam, who also participated in two years of prior design work with the team and, therefore had buy-in of the technology. Understandably, not all instructors used the orchestration tool in the same way or at the same frequency. For instance, Zain used the view work function more than the prompts. With the expectation of Zain, all instructors used the prompts more than the view work. Multiple avenues to interact with the orchestration tool are necessary to accommodate instructors' views and needs toward supporting students. Using actionable analytics that leverages data to inform concrete recommendations helps advise instructors on whom to help and why [8,7]. Nevertheless, designing the orchestration tool with multiple pathways to explore groups' progress and participation affords instructors the agency to make decisions without directing them on what to do and when [8,25]. Future research could explore how personalized training or customization of the tool could improve instructors orchestration strategies and how they support CPS.

All instructors were better at identifying *silent on-task* behaviors than group *off-task* ones. *Silent on task* is a relatively simple behavior to identify. *Off-task* becomes more challenging, as instructors can have

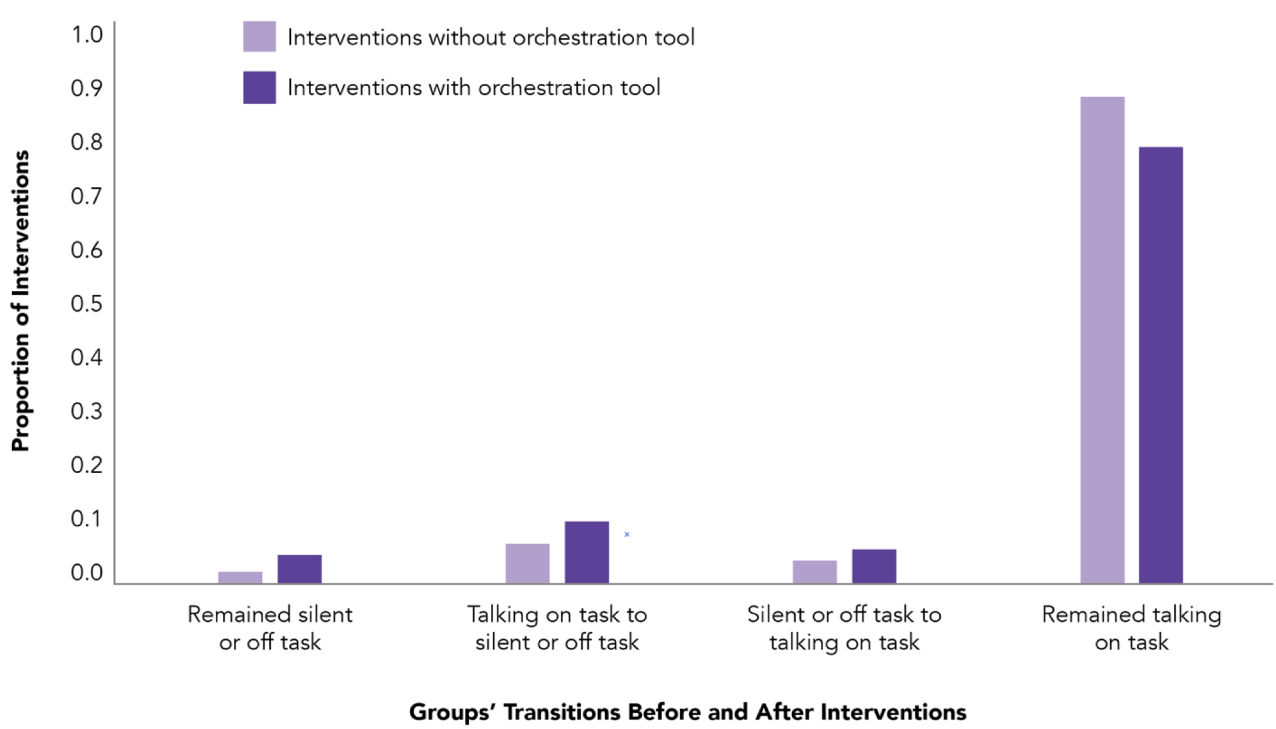


Fig. 8. Transitions across all interventions and with and without the orchestration tool.

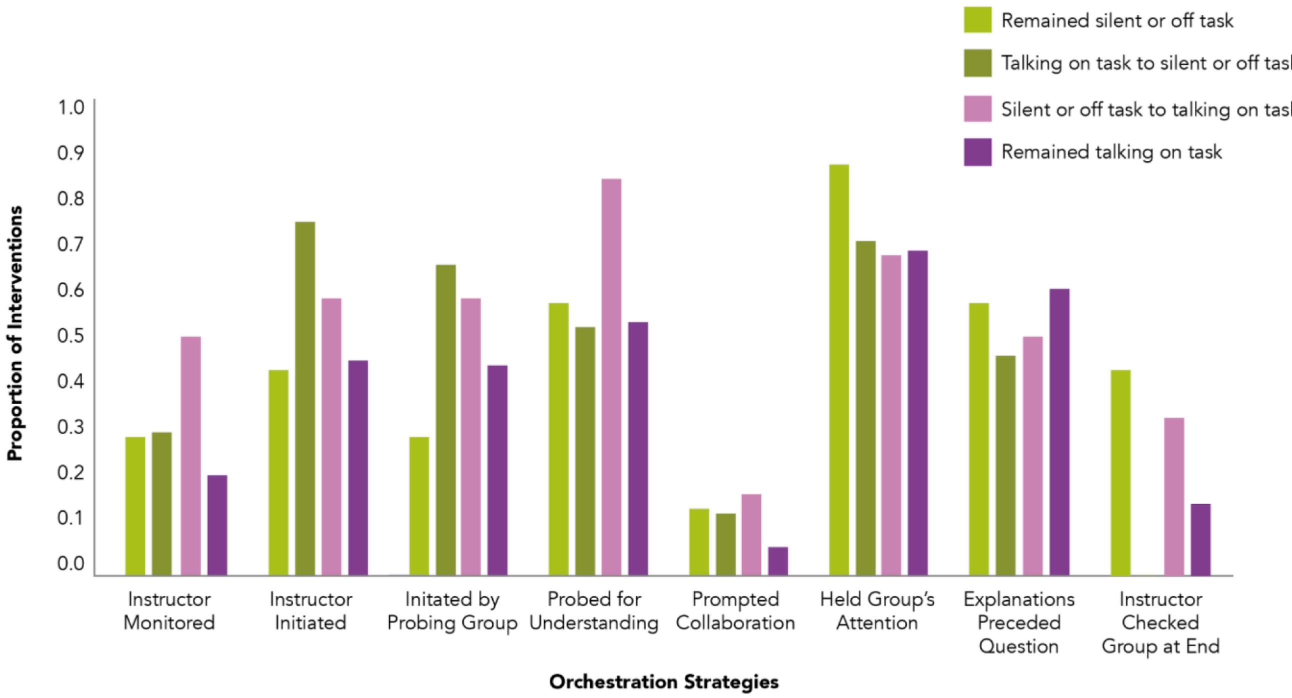


Fig. 9. Proportion of interventions by orchestration strategies categorized by transition.

different perceptions of what constitutes an *off-task* behavior. While we provided a definition in the tool, it was not representative of all possible *off-task* behaviors but rather a definition with an example, leaving it open to interpretation of what that means given the groups' current behavior. While the tool calls attention to specific moments and supports pedagogical decision-making [18], there are many scenarios the instructors may be responding to that extend beyond the support the tool gives. The *CSTEPS tool*, and others like it, must be responsive to various views and interpretations of predictive behaviors.

Lastly, examining the relationship between the model predictions, the expert and instructors' identification of behaviors, we found that instructors' identifications of *silent on-task* and *off-task* behaviors were largely correct. However, instructors correctly identified behaviors when the predictive models got them wrong. Instructors were much less likely to identify behaviors that the models got right. We attribute this to two main factors. First, since the models were mostly incorrect, it is probable that this led to instructors' doubt and lack of trust in the orchestration tool. While we decided to lower the thresholds to train instructors to recognize the predicted behaviors and to support possible moments where intervention might be helpful, through these decisions, we risk the instructors questioning the accuracy of the predictions. Second, because this is a classroom environment, many factors interfere with how instructors use the tool. For example, instructors responded to prompts often minutes after the prompt was delivered due to various reasons in the classroom, meaning the group's behavior may have changed after the model made predictions. To account for this, we designed our models to identify behaviors that persisted for at least a minute to capture sustained behavior. When creating orchestration tools to support CPS, we expect changes and tensions to arise in authentic classroom contexts. A goal of *CSTEPS tool* was to help instructors feel comfortable and confident supporting groups in spaces they are generally unprepared [11,14], we need to be able to support them without breaking their trust and accounting for the changes and disconnects that emerge in real-time.

Pedagogical findings and implications

Regarding pedagogical implications, we turn to our last two research questions. Our findings highlight evidence of potential interventions in real-time, machine learning-supported classrooms. The orchestration tool did impact some aspects of how instructors interacted with groups, specifically who prompted the intervention and how instructors started the conversation. There were no significant differences regarding groups' participation when instructors used the *CSTEPS tool*. Yet, we did see some promising results regarding orchestration strategies. When instructors used the tool and monitored before intervening, groups moved into productive participation. Additionally, there were moments when instructors applied prompts as expected, which led to effective interventions. While these were few, we see these as promising instances that can help us understand how to support instructors.

Looking closely at the use of specific orchestration strategies, we know that planning and facilitating CPS is a complex form of instruction that requires knowledge of this type of learning as well as strategies for monitoring and intervening in ways that promote collaboration among groups [5,6,11]. The contents of instructors' interventions are critical for productive group interactions [12], making pedagogical content knowledge an essential aspect of CPS classrooms. Monitoring is an essential step in gauging a group's progress through a CPS task and identifying their immediate needs [13], and we designed the orchestration tool intentionally to prompt instructors to monitor specific behaviors, share pointers on how to observe effectively, and share strategies for intervention. However, we saw no difference between monitoring behaviors with or without the orchestration tool, despite the tool directly prompting them to do so. Because we designed the technology to draw instructors to the groups, interventions without it are more likely to be prompted by groups. We observed that when using the tool, the instructors were more likely to initiate an intervention as compared to when they were not using it.

Regarding initiation moves, we found that instructors used less direct instruction and more open-ended questions and generative probes when using the tool. Instructors using less direct instruction is a promising result of the technology and the system change happening as part of our DBIR project. In our previous research in this context, engineering instructors often initiate interventions with instructions or the answer rather than probing and allowing the group to work together to explain what they are doing or need [6]. Prompting the group to work together was a core purpose of the orchestration tool [5,11]; however, instructors rarely used that strategy. We saw changes in the orchestration strategies of when instructors intervened and if and how they probed the group during the intervention. There is room to explore how we might support additional orchestration strategies in future iterations.

While some interventions moved groups from *talking and on task* to *silent* or *off-task*, such cases rarely occurred across the three weeks. While a few transitions before and after interventions were *silent* or *off-task* to *talking and on task*, we did see the frequency of some orchestration strategies (e.g., monitoring and probing for understanding) increase during those transitions, echoing the importance of these strategies in the literature findings [5,6].

This study provided many open design avenues for revising our tool to support instructors' orchestration strategies. First, future work can explore additional models and prompts that might effectively support instructors' orchestration strategies. Since CPS is complex, there are many aspects of students' learning processes to draw instructors' attention to foster productive social knowledge construction in CPS activities. Future studies need to construct machine learning models that help us understand a range of students' interactions and study their accuracy in authentic environments. In addition, we can think about how we might communicate models to provide more adaptive feedback to instructors. Adaptive feedback could help give instructors the support they need based on their intervention styles, experience with CPS, and comfort supporting groups. Our findings highlighted not all instructors leveraged the prompts due to many factors, such as the model's accuracy or factors in the classroom. Since no two instructors will have the same needs and groups engaged in CPS differ drastically based on their progress, interactions, and other factors, orchestration tools must be responsive to accommodate the dynamic versus static realities of classrooms. The *CSTEPS tool* had multiple modes of interacting with groups' participation, including real-time prompts and viewing students' work, but designers need to create multiple support that align with the dynamic elements of CPS classrooms.

Limitations

Our exploratory study looked at the accuracy of the machine learning models and how the instructors used the tool to support CPS. To do so we created models that predicted *off-task* and *silent on-task* behaviors, yet we recognized that these do not fully recognize the complexity of students' CPS and that we acknowledged we overpredicted these behaviors. We chose these behaviors because they were detectable through our work to create these machine learning models. We were interested in how prompting instructors to support these behaviors might impact them in context; however, we recognize that these behaviors paint a small, non-representative picture of what group CPS looks like [20]. By overpredicting these behaviors we may have lost the instructors trust in our tool; future work needs to predict other CPS-related behaviors and alter how to communicate them to support instructors and sustain trust. We also acknowledge this study is an exploratory study to see how machine learning models might impact orchestration strategies. Long term analyses are needed to understand the impact of the tool on orchestration strategies and CSP behaviors. Lastly, since we worked with a small number of instructors, with one in particular who worked with us on the research team, this may have impacted their use of the tool. We highlight the need for other methods to better understand the accuracy of models, such as setting up a time gap in the instructors' procedure or iterative implementation cycles to understand how instructor interventions changed.

Conclusion

Our paper contributes to the evolving field of CPS and orchestration by offering insights into the interplay between predictive models, instructional tools, and orchestration strategies in engineering classrooms. The findings underscore the nuanced nature of classroom orchestration and emphasize the potential of technology to enhance real-time interventions and support collaborative participation. Our paper makes three main contributions to the literature. First, our study illustrates the development and testing of a novel, real-time machine learning model and prompts for supporting instructors' interventions. This work moves beyond monitoring and mirroring orchestration support and provides empirical evidence of interventions to support CPS. Second, we provide evidence of potential interventions in machine-learning supported classrooms. Our findings show how instructors respond to and use real-time, machine-learning strategies to support CPS classrooms. Lastly, we provide insights into a DBIR approach and a detailed analytic process for triangulating data types to understand the tool's impact.

CRediT authorship contribution statement

LuEttaMae Lawrence: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emma Mercier:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Taylor Tucker Parks:** Writing – review & editing, Validation, Formal analysis, Data curation. **Nigel Bosch:** Writing – review & editing, Software, Resources, Project administration, Investigation, Data curation. **Luc Paquette:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Example task

Name: ______________

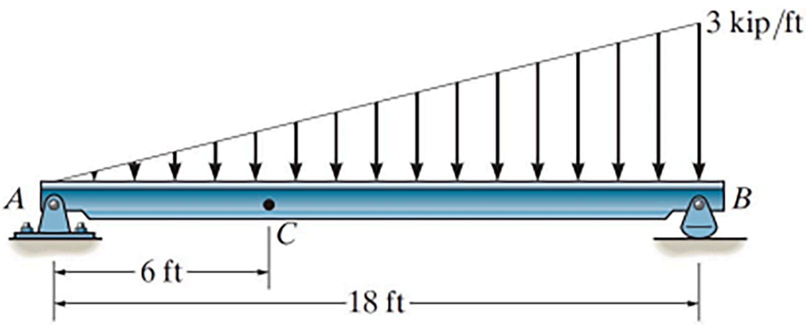
Group members: ____________________ ____________________ ____________________

TAM 210/211 - Worksheet 10

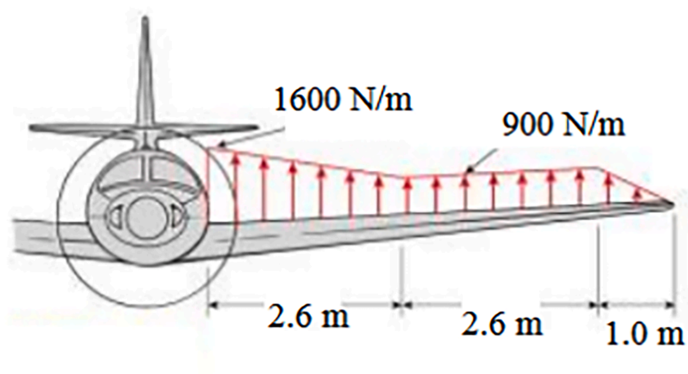
1) For the entire beam below

a) obtain the internal shear $V(x)$ and bending moment $M(x)$

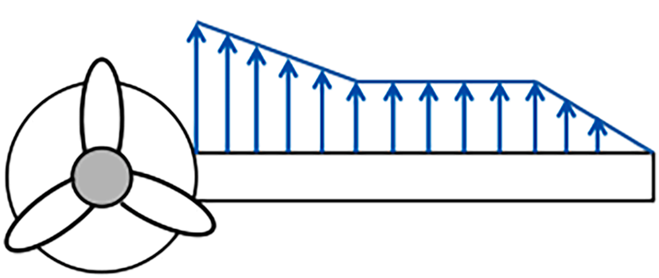
b) draw the shear and bending moment diagrams



2) Under cruising conditions the distributed load acting on the wing of a small airplane has the idealized variation illustrated below.



Below is a simplified version of the wing loading



a) determine the reactions at the inboard end of the wing

Appendix B. Coding Scheme

Code	Code	Description	IRR
Instructor monitored	1 = yes	Instructor monitored the group for at least 10 s	Agree: 97 % Kappa: 0.93
	0 = no	Instructor monitored the group for less than 10 s or not at all	
Instructor initiated	1 = instructor	Instructor intervened first	Agree: 96 % Kappa: 0.92
	0 = student	Student asked a question	
Initiated by probing group	1 = prompting conversation	Instructor began the intervention by asking the group how/what they were doing, if they were stuck, or prompting them to collaborate (e.g., what did you get?, what are you struggling with?, can you tell him how you got there)	Agree: 99 % Kappa: 0.96
	0 = delivering instruction or directions	Instructor began the intervention by instructing the group what to do or telling them the answer (e.g., No, that answer is 5; That works but would be a big cross section, right?)	
Prompted Collaboration	1 = yes	The instructor explicitly prompt the group to interact with each other (e.g., encouraging the group to talk, respond to each others' questions, noting they're on different pages, or collaborate on the task)	Agree: 93 % Kappa: 0.25
	0 = no	The instructor does not explicitly prompt the group to interact	
Probed for understanding	1 = yes	Occurance of any questions about the groups' understanding, what they are doing, or if they are stuck that took place after the initiation move	Agree: 89 % Kappa: 0.77
	0 = no	All turns after the initiation move were instruction where the instructor was telling the group what to do, explaining a concept, or providing an answer	
Held group's attention	1 = yes	All students were engaged, speaking with, or listening to the instructor during the intervention or the instructor is attempting talk to all group members	Agree: 92 % Kappa: 0.87
	0 = no	The instructor did not have the entire groups full attention (e.g., only spoke to one or two students)	
Explanations proceeded quetsions	1 = yes	Instructor provided elaborated description because a student asked a question or expressed they were confused	Agree: 93 % Kappa: 0.86
	0 = no	Instructor provided elaborated description without prompting from students	
Instructor checked group at the end	1 = yes	Instructor ended the intervention by asking if the students understood or if they had additional questions	Agree: 99 % Kappa: 0.95
	0 = no	Instructor did not check with students, but ended the intervention in any other way	

References

[1] Rummel N, Spada H. Learning to collaborate: an instructional approach to promoting collaborative problem-solving in computer-mediated settings. J Learn Sci 2005;14(2):201–41.

[2] Hesse F, Care E, Buder J, Sassenberg K, Griffin P. A framework for teachable collaborative problem-solving skills. In: Care E, Griffin P, Wilson M, editors. Assessment and teaching of 21st century skills. Springer; 2015. p. 37–56.

[3] Roschelle J, Teasley SD. The construction of shared knowledge in collaborative problem-solving. In: O'Malley C, editor. (editor), Computer supported collaborative learning. Springer; 1995. p. 69–97.

[4] Vygotsky LS. Mind in society: development of higher psychological processes. Harvard university press; 1978.

[5] Kaendler C, Wiedmann M, Rummel N, Spada H. Teacher competencies for the implementation of collaborative learning in the classroom: a framework and research review. Educ Psychol Rev 2015;27(3):505–36. https://doi.org/10.1007/s10648-014-9288-9.

[6] Shehab SS. Collaborative problem-solving in higher education classrooms: Exploring student interactions, group progress, and the role of the teacher. University of Illinois at Urbana-Champaign; 2019.

[7] Vieira C, Parsons P, Byrd V. Visual learning analytics of educational data: a systematic literature review and research agenda. Comput Educ 2018;122:119–35.

[8] van Leeuwen A, Rummel N, Van Gog T. What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? Int J Comput Support Collab Learn 2019;14(3):261–89. https://doi.org/10.1007/s11412-019-09299-x.

[9] Sun C, Shute VJ, Stewart A, Yonehiro J, Duran N, D'Mello S. Towards a generalized competency model of collaborative problem-solving. Comput Educ 2020;143:103672. https://doi.org/10.1016/j.compedu.2019.103672.

[10] Nelson LM. Collaborative problem-solving. Instruct Des Theor Models 1999;2(1999):241–67.

[11] Webb NM. The teacher's role in promoting collaborative dialogue in the classroom. Br J Educ Psychol 2009;79(1):1–28. https://doi.org/10.1348/000709908x380772.

[12] Webb NM, Franke ML, Ing M, Wong J, Fernandez CH, Shin N, Turrou AC. Engaging with others' mathematical ideas: interrelationships among student participation, teachers' instructional practices, and learning. Int J Educ Res 2014;63:79–93.

[13] Kaendler C, Wiedmann M, Leuders T, Rummel N, Spada H. Monitoring student interaction during collaborative learning: design and evaluation of a training program for pre-service teachers. Psychol Learn Teach 2016;15(1):44–64. https://doi.org/10.1177/1475725716638010.

[14] Lawrence L. University of Illinois at Urbana-Champaign. The design process of a collaborative orchestration tool and its implications for instructor uptake (Doctoral dissertation. 2020.

[15] ABET. (2018, April). Criteria for accrediting engineering programs (for 2018-2019). ABET. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2018-2019/.

[16] Dillenbourg P, Jermann P. Technology for classroom orchestration. In: Khine MS, Saleh IM, editors. New science of learning. Springer; 2010. p. 525–52.

[17] Roschelle J, Dimitriadis Y, Hoppe U. Classroom orchestration: synthesis. Comput Educ 2013;69:523–6. https://doi.org/10.1016/j.compedu.2013.04.010.

[18] Song Y. A review of how class orchestration with technology has been conducted for pedagogical practices. Educ Technol Res Dev 2021;69:1–27. https://doi.org/10.1007/s11423-021-10001-y.

[19] Tarmazdi H, Vivian R, Szabo C, Falkner K, Falkner N. Using learning analytics to visualise computer science teamwork. In: Proceedings of the 2015 ACM Conference on Innovation and technology in computer science education; 2015. p. 165–70.

[20] Van Leeuwen A, Janssen J. A systematic review of teacher guidance during collaborative learning in primary and secondary education. Educ Res Rev 2015;27:71–89.

[21] Perera D, Kay J, Koprinska I, Yacef K, Zaïane OR. Clustering and sequential pattern mining of online collaborative learning data. IEEE Trans Knowl Data Eng 2008;21(6):759–72.

[22] Martinez-Maldonado R, Gaševic D, Echeverria V, Fernandez Nieto G, Swiecki Z, Buckingham Shum S. What do you mean by collaboration analytics? A conceptual model. J Learn Anal 2021;8(1):126–53.

[23] Dimitriadis Y, Martínez-Maldonado R, Wiley K. Human-centered design principles for actionable learning analytics. Research on E-Learning and ICT in education. Cham: Springer; 2021. p. 277–96.

[24] Jermann P, Soller A, Muehlenbrock M. From mirroring to guiding: a review of the state of art technology for supporting collaborative learning. In: European Perspectives on Computer-Supported Collaborative Learning; 2001. p. 324–31. No. CONF.

[25] An P, Holstein K, d'Anjou B, Eggen B, Bakker S. The TA Framework: designing real-time teaching augmentation for K-12 classrooms. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; 2020. p. 1–17.

[26] Lawrence, L., Rummel, N., & Aleven, V. (2022). Ethical consideration for designing AI to support dynamic learning transitions.

[27] Fishman BJ, Penuel WR, Allen AR, Cheng BH, Sabelli NORA. Design-based implementation research: An emerging model for transforming the relationship of research and practice. Teach Coll Rec 2013;115(14):136–56.

[28] Shehab S, Mercier E, Kersh M, Juarez G, Zhao H. Designing engineering tasks for collaborative problem solving. In Making a Difference—Prioritizing Equity and Access in CSCL: The 12th International Conference on Computer Supported Collaborative Learning. Philadelphia: The International Society of the Learning Sciences 2017.

[29] Shehab S, Lawrence L, Mercier E, Margotta AS, Livingston ER, Silva M, Tucker T. Towards the effective implementation of collaborative problem solving in

undergraduate engineering classrooms: co-designing guidelines for teaching assistants. American Society for Engineering Education; 2020.
[30] Tucker T, Lawrence L, Mercier E. Investigating the Effectiveness of an Orchestration Tool on the Nature of Students. American Society for Engineering Education; 2021.
[31] Lawrence L, Mercier E. Co-Design of an Orchestration Tool: Supporting Engineering Teaching Assistants as they Facilitate Collaborative Learning. Interact Des Archit (s) 2019;42:111.
[32] Hur P, Bosch N, Paquette L, Mercier E. Harbingers of collaboration? The role of early-class behaviors in predicting collaborative problem-solving. In: Rafferty AN, Whitehill J, Cavalli-Sforza V, Romero C, editors. Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020); 2020. p. 104–14.
[33] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[34] Greene JC. Toward a methodology of mixed methods social inquiry. Res Schools 2006;13(1):93–8.
[35] Sullivan C, Forrester MA, editors. Doing qualitative research in psychology: A practical guide. Sage; 2018.