1 INTRODUCTION

With the rapid emergence and adoption of learning technologies today, students are gaining more autonomy and independence in their education than ever before. As such, it becomes increasingly important that the teacher's role is supported and reinforced through classroom orchestration tools.

Current orchestration tools exist for individual learning activities [1] and collaborative learning activities [2]; however, there is little research on orchestrating dynamic transitions between individual and collaborative activities in the classroom. Dynamically switching between individual and collaborative activities allows students to experience a variety of instruction catered to their specific needs. For teachers, managing these dynamic transitions is complex and time-consuming. The teacher must plan, manage, and monitor students’ learning activities and processes in real time.

Previous work on this topic suggests a combination of human and AI control, where various agents in the classroom, including the teacher, the students, and the AI system, share in the classroom orchestration [3]. This allows the teacher to let the system and/or students have a some form of control over the classroom by making pairing decisions themselves, lessening the burden on the teacher. Yet, it remains unclear what exact degree of control would be desirable, versus overburdening or distracting, to teachers in these contexts.

In particular, we focus on the collaborative activity of peer tutoring, where one student, the tutee, learns from another student, the tutor. In a peer tutoring activity, both the tutor and tutee can mutually benefit. The tutee is able to see the problem through the lens of the tutor, while the tutor enhances his understanding of the material.

With the goal of developing a teacher orchestration tool for dynamic transitions between individual and peer tutoring activities in face-to-face classrooms, two issues are explored: how to determine when peer tutoring is more beneficial than working individually, and how to pair students effectively.

Following a technology probed approach, a classroom pilot of APTA, an AI-based system designed to help students engage in peer tutoring, is conducted in which teacher-driven, student-driven, and system-driven pairing policies are compared. Triggers and parameters for system suggested pairings are developed, and a prototype of a dashboard for teachers
to monitor these system suggestions is created and shown to teachers in a design workshop.

2 METHODS

In order to understand how the orchestration tool could best be designed, we used two AI-based tutoring systems: Lynnette and APTA 2.0, as support tools in our studies. We conducted two studies: a classroom pilot and a design workshop.

2.1 Learning Context

Both Lynette and APTA support linear equation solving activities at the middle school level. Lynette is a rule-based cognitive tutor [4] within the CTAT/Tutorshop architecture [5]. It supports individual problem solving activities through hints, correctness verifications, and skill level updates. A screenshot of Lynnette is shown in Figure 1.

APTA [6], which stands for Adaptive Peer Tutoring Assistant, is an AI-based system built on top of Lynnette designed to help students engage in peer tutoring. We reimplemented APTA, creating APTA 2.0, to support two different interfaces: one for the tutor, and the other for the tutee. This allows the tutee and tutor work synchronously across separate devices.

A screenshot of the tutee’s screen in APTA 2.0 is displayed in Figure 2, and a screenshot of the tutor’s screen in APTA 2.0 is displayed in Figure 3. The tutee’s role is to solve the problem, and his work is synchronously updated and displayed on the tutor’s screen. The tutor’s job is to grade the tutee and provide guidance to the tutee.

In APTA 2.0, the tutor has access to hints through the Lynnette model, but the tutee does not. As such, the tutee
must ask the tutor for hints if he needs guidance. Communication between the tutor and tutee is accomplished through a messaging interface embedded within APTA 2.0. A coaching model monitors the peer tutoring progress and intervenes in the messaging system as necessary, as can be seen in Figure 2 and Figure 3. The tutee cannot move on until the tutor has agreed that the tutee is correct through his grading.
2.2 Classroom Pilot

We conducted a preliminary classroom pilot at a middle school in the Pittsburgh region. A total of 118 students in six seventh grade math classes, across three different teachers, were involved. One of the teachers was female, and the other two were male. Two of the teachers had previously participated in the design of orchestration tools [9].

Students began working individually using Lynnette. After 15 minutes, a randomly assigned pairing policy was enacted through which some students were paired and began working with APTA for the remainder of the class period. The pairing policies were as follows:

- **Student driven**: Students request to be paired by providing a list of classmates (potential tutors) that they would like to work with in the order of their preference. From this list, a student who is “free” (not already in a pair) is chosen to be the student’s tutor, with the higher preference choices prioritized.

- **Teacher driven**: Teachers pair students by identifying struggling students (tutee) and students who could help them (tutor) on a case by case basis, and the corresponding pairs are created.

- **System driven**: A researcher (simulating an AI system) monitors the students using Tutorshop [5], pairing them based on skill level. Students with a skill level below 50% for any particular skill are paired with students with a skill level above 75% for the same skill, with the lower skill level student as the tutee and the higher skill level student as the tutor.

Following this session, the teacher led the class in a workshop session. Students were asked about their own preferences regarding individual and collaborative activities, as well as their thoughts on how helpful and effective the pairing policy they experienced was. They were also given a verbal description of the other two pairing policies they were not able to experience and asked to compare and contrast them with what they experienced.

After all class sessions were complete, the teachers participated in a semi-structured interview. They were asked about their control preferences for each of the pairing policies and which policy they best preferred. Furthermore, teachers were asked about hybrid cases, where two or all of the pairing policies were combined, and whether the combination would be even more beneficial than a single pairing policy.

Following the conclusion of the classroom pilot, the feedback from the pilot was summarized, and design ideas formed for an orchestration tool for dynamically transitioning between individual and peer tutoring activities. This led to the development of several interactive prototypes of dashboards for the teacher to control and monitor pairings.

2.3 Design Workshop

The dashboard prototype was then reviewed by two teachers in a design workshop, conducted remotely over a joint video call. Both teachers were male middle school math teachers. One of the teachers was a participant from the preliminary classroom pilot, while the other was from the Chicago region and had previous experience with Lynnette.

Teachers were asked to interact with the prototypes and think out loud while interacting. Questions regarding their
thoughts on how “in control” the prototype made them feel and their trust of the dashboard system were asked.

3 RESULTS

3.1 Classroom Pilot

From the student workshops conducted during the classroom pilot, students expressed various opinions regarding their preferences for individual and collaborative work. A common opinion expressed was that they preferred working on most problems by themselves so that they could go at their own pace; however, when faced with more challenging and difficult problems, they preferred working collaboratively.

Regarding the pairing policies, some students liked being able to pick their own partners, but many stated that the teacher would pick a more suitable partner for them. Many stated that it was “exciting” when either the teacher or the system paired them because they did not know who their partner would be.

During the classroom pilot teacher interviews, teachers generally stated that with regards to the pairing policies tested, their comfort with student driven pairings depended on the specific students asking for help and the specific class dynamics. They found that teacher driven pairing was extremely time-consuming and stressful. This was mainly because pairs were created dynamically, which meant that teachers had to spend most of the class time reasoning about and creating pairs rather than actually helping students with the conceptual knowledge of how to solve the problems. Teachers felt that they could better spend their class time by helping students directly.

As such, teachers preferred the hybrid approach of system driven pairing with the added ability for them to accept or reject the pairs created by the system. Teachers stated that it was important for them to have the power to override system driven pairing by verifying the system suggested pairs, because sometimes their own judgement would be more accurate than the system’s. Some students simply could not work well together, and the teacher would have to intervene during system driven pairing in order to prevent those specific students from being paired.

The data collected regarding the pairing frequencies for each pairing policy enacted, as shown in Table 1, indicates that the student driven pairing policy resulted in the smallest fraction of the students paired. On the other hand, the teacher and system driven pairing policies both led to a larger fraction of the students paired. The teacher and system driven pairing policies paired about the same fraction of the students, suggesting that both the teacher and system were equally active in creating pairing instances. Furthermore, this indicates that student driven pairing might be best supported by either teacher driven or system driven pairing, or a combination of both.

Table 2 shows the average error rates for each class, comparing the average error rate when working individually to the average error rate when working in pairs. Across all classes, average error rates when working individually were smaller than average error rates when working in pairs. This may have been attributed to the fact that students were generally more focused when working individually, given that many of them had prior experience working with similar individual learning activities. Once the pairing policies were
Table 1: Pairing frequency per pairing policy

<table>
<thead>
<tr>
<th>Pairing policy</th>
<th>Pairing frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>0.195</td>
</tr>
<tr>
<td>Teacher</td>
<td>0.419</td>
</tr>
<tr>
<td>AI System</td>
<td>0.413</td>
</tr>
</tbody>
</table>

Table 2: Average error rates per class

<table>
<thead>
<tr>
<th>Class</th>
<th>Pairing Policy</th>
<th>Individual</th>
<th>Peer Tutoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>0.02 ± 0.32</td>
<td>0.24 ± 0.76</td>
</tr>
<tr>
<td>2</td>
<td>Student</td>
<td>0.17 ± 2.98</td>
<td>17.18 ± 79.81</td>
</tr>
<tr>
<td>3</td>
<td>Teacher</td>
<td>0.15 ± 1.11</td>
<td>1.09 ± 2.95</td>
</tr>
<tr>
<td>4</td>
<td>Teacher</td>
<td>0.18 ± 0.85</td>
<td>1.05 ± 3.04</td>
</tr>
<tr>
<td>5</td>
<td>System</td>
<td>0.21 ± 2.37</td>
<td>2.18 ± 5.68</td>
</tr>
<tr>
<td>6</td>
<td>System</td>
<td>0.41 ± 2.06</td>
<td>0.43 ± 1.24</td>
</tr>
</tbody>
</table>

introduced, students became more confused and unfocused, because the majority had not been exposed to a peer tutoring system before.

By matching each class to the pairing policy that was assigned to it, we were able to compare the average peer tutoring error rate across the different pairing policies in Table 2, giving us an idea of how effective each pairing policy was.

As shown, results for student driven pairing were incredibly variable between the two classes where it was tested. This supports the teachers’ comments that the success of student driven pairing highly depends on the specific students and class dynamics. On the other hand, teacher and system driven pairing policies led to error rates that were about the same. The teacher driven pairing led to overall more consistent results across the classes, which also supports their comments that at times, their own judgement would be better than the system’s judgement for pairings.

3.2 Dashboard Prototypes

The screenshots of the dashboard prototypes developed, using the feedback from the teachers and students in the classroom pilot, are shown below. Figure 4 shows an example of the dashboard when the system suggests a pair, Figure 5 shows an example of when the student suggests a pair, and Figure 6 shows an example of when the teacher creates a pair.

A dashboard integrating all of these prototypes together allows the teacher to have ultimate control of the pairings by pushing the system suggested and student suggested pairings to the teacher for verification, and allowing the teacher to create his/her own pairings as appropriate.

All three dashboard prototypes contain information about the current classroom, as displayed by the small symbols and shapes around certain students’ names. Currently paired students have an oval around their name with a color unique to their pair, and the tutor of each pair is marked by the small teacher image above the their name. A red question
Figure 4: Prototype 1 - System suggested pairing

Figure 5: Prototype 2 - Student suggested pairing
mark above a student’s name indicates that the student is struggling, which is defined by many attempts to solve a problem and little progress made. This notion of struggling can also be referred to as wheel-spinning [7]. A caution sign above a student’s name indicates that the student is “gaming the system” [8], which is defined by abuse of hints and the tutoring system in order to find a shortcut to the answer.

3.3 Design Workshop

While teachers were interacting with prototype 1 for system suggested pairings, they agreed that they wanted detailed information about the reasoning behind each system suggested pairing. This additional information would better inform their decision to either accept or reject the particular proposed pairing.

For prototype 2 for student suggested pairings, teachers agreed that whether they would accept or reject the student
proposed pairing would vary on a case by case basis. This is similar to the responses that we received from the classroom pilot regarding the student driven pairing policy.

While interacting with prototype 3 for teacher driven pairings, teachers stated that they strongly preferred to choose from a list of system suggested pairings rather than manually from a list of the available students who were currently working individually. They agreed that a manual selection purely from the available students would be necessary in some cases, for example, when the system does not have enough information about a particular student to generate suggested tutors for him/her. However, they said that the manual selection option should only be present if needed.

Furthermore, teachers were interested in having a pairing history of the effectiveness of past tutoring sessions, which could help inform pairing decisions. This pairing history would be able to capture how well particular pairs of students worked together. Teachers also noted that in this prototype, actions were taken to pair a student who was gaming the system. They both noted that while having an indication of students who were gaming the system was useful, they would not treat it as a reason to put those students in pairs. Instead, they would prefer to directly intervene by speaking one-on-one with the student.

4 LESSONS LEARNED

Through this project, we learned that there is definitely a need for a hybrid form of orchestration control in order to facilitate dynamic transitions between individual and peer tutoring activities. Teachers want to be in control of the classroom by verifying tutor-tutee pairs, but do not necessarily want to have the responsibility of creating the pairs. They generally seem hesitant about students creating their own pairs because of the high variability in effectiveness of the pairs, which we could also see from the data collected during the classroom pilot. This leads to a hybrid form of control and orchestration purely between the teacher and the system.

Furthermore, teachers want to be aware of the actions occurring in the orchestration tool. They want to know the reasoning behind the pairs suggested by the system as well as successful pairing history, and be able to utilize them to create the most effective tutor-tutee pairs.

Additionally, it is extremely important that this orchestration tool is able to adapt to specific preferences of the teacher. Since every teacher is different, one teacher may judge the effectiveness of a tutor-tutee pair differently from another. Because we want to give the teacher the highest level of control in this orchestration system, the orchestration tool must respect the differences of each teacher. Thus, the orchestration tool must allow teachers to configure the pairing system to their liking.

Through our work with students and teachers, we propose that system suggested pairings are triggered by indication of struggle, which can be implemented by wheel-spinning [7] detectors embedded in Lynnette. The criteria for the system suggested tutor-tutee pairings is history and skill level.

The history criterion used to pair students is based on successful previous pairings, where success is defined as the tutee’s improvement combined with the tutor’s helpfulness. The tutee’s improvement can be measured by the change in
error rate and skill level over the time that he is being peer tutored, while the tutor’s helpfulness can be measured by classifications produced by the NLP model embedded in the APTA messaging system.

On the other hand, the skill level criterion used to pair students was debated by various teachers. Some teachers indicated in the design workshop that they wanted to pair students with the same skill level, while others saw contrasting differences in skill level as a stronger reason for pairing. This difference in teacher opinion points to the future development of a configuration step for teachers to define their specific preferences regarding the skill level criterion.

5 CONCLUSIONS

Based on the feedback from teachers and students obtained through the classroom pilot and design workshop, we suggest an adaptive orchestration tool which provides the teacher with the ultimate control of dynamic transitions between individual and peer tutoring activities.

While the teachers are in ultimate control, they rely on the additional agent of the system to support their role. The system can propose suggested tutor-tutee pairings to the teacher. This allows the teacher to simply accept or reject the pairing proposals instead of spending their valuable and limited class time attempting to create the most effective peer tutoring pairs purely from their own judgement.

6 FUTURE WORK

We would like to generalize these findings by having more middle school math teachers, we would also like to confirm these findings with teachers teaching different subjects and different grade levels. This means that the types of problems that APTA is able to support must expand past the current linear equations model and to other subject and problem types. In turn, this will allow the orchestration system to be used in a much broader scope of the teacher’s curriculum.

We would also like to implement and test the design ideas represented by the classroom dashboard prototype by creating the actual web browser-based dashboard. This includes implementing our proposed system pairing AI and performing more classroom pilots in order to fine tune the system.

Additionally, we would like to develop and test a configuration step in the orchestration tool for teachers to define their specific preferences regarding the system suggested pairings, allowing them to configure the pairing system to their preferences.

We hope that the implementation of this orchestration tool will be able to assist the teacher in utilizing system suggested pairings to create an environment in which tutors and tutees mutually benefit.

7 REFERENCES


