A Spatiotemporal Analysis of Teacher Practices in Supporting Student Learning and Engagement in an AI-enabled Classroom

Shamya Karumbaiah[1], Conrad Borchers[2], Tianze Shou[2], Ann-Christin Falhs[3], Pinyang Liu[2], Tomohiro Nagashima[4], Nikol Rummel[3], Vincent Aleven[2]

[1]University of Wisconsin-Madison, Madison, USA
[2]Carnegie Mellon University, Pittsburgh, USA
[3]Ruhr University Bochum, Bochum, Germany
[4]University of Saarland, Saarbrücken, Germany

shamya.karumbaiah@wisc.edu, cborcher@cs.cmu.edu, tshou@andrew.cmu.edu, ann-christin.falhs@rub.de, pinyangl@andrew.cmu.edu, nagashima@cs.uni-saarland.de, nikol.rummel@rub.de, vincent.aleven@cs.cmu.edu

Abstract. Research indicates that teachers play an active and important role in classrooms with AI tutors. Yet, our scientific understanding of the way teacher practices around AI tutors mediate student learning is far from complete. In this paper, we investigate spatiotemporal factors of student-teacher interactions by analyzing student engagement and learning with an AI tutor ahead of teacher visits (defined as episodes of a teacher being in close physical proximity to a student) and immediately following teacher visits. To conduct such integrated, temporal analysis around the moments when teachers visit students, we collect fine-grained, time-synchronized data on teacher positions in the physical classroom and student interactions with the AI tutor. Our case study in a K12 math classroom with a veteran math teacher provides some indications on factors that might affect a teacher's decision to allocate their limited classroom time to their students and what effects these interactions have on students. For instance, teacher visits were associated more with students' in-the-moment behavioral indicators (e.g., idleness) than a broader, static measure of student needs such as low prior knowledge. While teacher visits were often associated with positive changes in student behavior afterward (e.g., decreased idleness), there could be a potential mismatch between students visited by the teacher and who may have needed it more at that time (e.g., students who were disengaged for much longer). Overall, our findings indicate that teacher visits may yield immediate benefits for students but also that it is challenging for teachers to meet all needs - suggesting the need for better tool support.

Keywords: Spatial analytics, Temporality, Teaching, Student Engagement, Human-AI Partnership, Multimodality.

1 Introduction

Previous studies suggest that teachers play an active role in supporting student learning with AI tutors [10,2]. With adaptive instruction and immediate feedback, it is often argued that AI tutors are designed to free up teachers' time so they can focus on students who need their time the most [9]. Accordingly, teachers are observed to be highly engaged in such AI-enabled classrooms, moving from one student to another providing individualized help [2] and socio-emotional support [11], and helping students get out of an unproductive rut [9]. Yet, our scientific understanding of teacher practices around AI tutors in classrooms and the ways in which they mediate student learning is limited [cf. 9,17]. Besides being interesting in its own right, this kind of understanding could be helpful in designing better tools that support teachers as they help students during AI-enabled classroom sessions, a focus within AIED research in recent years [9,13].

1.1 Spatiotemporal Factors in Teacher Practices with AI Tutors

To better understand teacher practices in classrooms with AI tutors, we need to go beyond student log data (an often-used source of data in AIED research) and explore the physical classroom where the student-teacher interaction is happening while students are learning with AI tutors [7]. In a previous study, Holstein and colleagues [10] highlighted the importance of analyzing spatial factors of student and teacher behaviors to understand the role of "out-of-software events" such as teachers' help-giving in AI-enabled classrooms. More broadly, research in education highlights the importance of studying teachers' and students' physical location and interaction in a classroom to understand teachers' pedagogical practices and their impact on students [5,16]. More recently, spatial pedagogy [14] - a framework on teacher positioning and movement in classrooms - has been used in learning analytics to better understand teachers' spatial pedagogical approaches [15]. However, it is hard to borrow these insights as-is for AI-enabled classrooms since the presence of the AI tutor drastically transforms the classroom environment and its social structures [2]. For example, the teacher's role shifts from that of a lecturer at the front of the class to a coach working with students individually in their close proximity [2,9].

Due to the often one-to-one nature of teachers' individualized attention to students in AI-enabled classrooms [2], we focus our spatial analysis on teachers' close physical proximity to individual students. We use the term *teacher visits* to denote shorter or longer episodes where the teacher stays close to the same student (see illustration in Figure 1; *left*). Teacher visits to students in AI-enabled classrooms could serve several purposes, such as monitoring, help-giving, motivational support, or simple reassurance. Teachers may visit an individual student or a small group of students after observing them from afar or while monitoring them routinely. Often, these visits help them gather additional information to improve their sensemaking and interpretation of noteworthy events (aka shaping [17]). Some teacher visits may involve specific intervention (e.g., conceptual guidance, socio-emotional support), either proactively based on what they notice or find out through talking to students directly, or after being prompted by a student's request for help (e.g., by raising their hand). While teacher visits have been reported to be positively related to learning and engagement in traditional classroom settings [6], they have not been studied thoroughly in human-AI hybrid teaching. One notable exception is a study by Holstein et al. [10] which collected manual observations of teacher visits and found a positive effect of teacher monitoring on student learning.

While exploring the spatial dimension of student-teacher interactions in classrooms with AI tutors seems promising, we argue that it is necessary to further contextualize the role of teacher visits by juxtaposing it with student learning and engagement happening in the AI tutor just prior to and just after each visit. That is, instead of aggregating measures of students' learning or engagement, we look at change (e.g., in idleness or struggle [12]) from before to after a teacher visit. Doing so enables us to explore factors that may correspond to a teacher's choice to visit a given student at a given point in time, and what change each teacher visit may bring for the visited student. This quantitative spatiotemporal lens complements earlier qualitative explorations of teacher practices using a replay tool [10].

1.2 Research Questions and Hypotheses

Prior interviews of teachers in AI-enabled classrooms (e.g., [10]) help us identify two key factors that may help determine a teacher's choice of the student to visit: 1) teachers' prior knowledge about student abilities and behavioral tendency, and 2) their perception of student needs while observing the class and monitoring student work on the students' computer screens. For our analysis, we translate these into two types of proxies: 1) a broader, static measure of student need for teacher support such as low prior knowledge, and 2) students' in-the-moment struggle and disengagement while learning with an AI tutor. Determining students' in-the-moment needs by direct observation may be harder for teachers in classrooms with AI tutors since there is a gap between what the teacher can perceive in the physical classroom and what their students do behind the screens with the AI tutor [9]. Certain student disengagement behaviors may be more overt (e.g., being idle) than others (e.g., misusing the tutor) for teachers' direct observation. A previous qualitative analysis showed that despite wanting to pay attention to students who are off-task, teachers overlooked students who spent greater time off-task [10]. Also, more so than in traditional classrooms, students are likely to work on divergent activities in the AI tutor and have different needs for teacher conceptual guidance while struggling, which makes us ask if there are differences in the effectiveness of teacher intervention based on student needs. Lastly, since teacher time is limited, we also expect teachers' decisions on who they will visit to depend on what they perceive a student's need is relative to other students' needs at a given time. These motivations translate to the following research questions and hypotheses:

RQ1. What factors about student learning and engagement with AI tutors relate to a teacher's choice of students they visit?

H1: Teacher visits are related to broader, static measures of student need for teacher support (i.e., low prior knowledge).

H2a: Teacher visits are related to students' in-the-moment needs while learning with the AI tutor, that is, students who are currently struggling or disengaged (i.e., being idle, misusing the tutor).

H2b: Teacher prioritizes students who have been struggling or disengaged the longest.

RQ2. How do teacher visits relate to student learning and engagement with AI tutors?

H3: Teacher visits are associated with less struggle or disengagement after the visit has taken place, compared to before.

H4: Teacher visits positively relate to learning, as measured by in-tutor performance and out-of-tutor knowledge tests.

Answering these questions could generate insights on analytics that could aid teachers while co-orchestrating their classrooms with AI tutors and generate scientific knowledge about effective teaching practices that, in turn, can support teacher learning and reflections. To enable such investigations, we need time-synchronized data about teachers' visits and student interaction with the AI tutor. In this work, we use position sensors to automatically record teachers' positions in the classroom and algorithmically infer visits (with reasonable albeit imperfect accuracy; [8]). We detail our methodology in the next section. Then, we present the findings from a case study in an authentic setting (i.e., a K12 math classroom using an AI tutor), addressing the research questions above. Lastly, we discuss the methodological and empirical contributions of this work on understanding teacher practices in supporting student learning with AI tutors.

2 Methods

2.1 Case Study Context

For the duration of three days in Summer 2022, we collected teacher position and student tutor log data in a public school in the United States. The participants were eighty-five 7th graders across five different classes (aka Periods 1-5), all taught by the same mathematics teacher. At the school, in 2022, 45.9% of all students were categorized as "Below Basic" for their performance on the end-of-course test on Algebra 1 [3]. The teacher who participated in the study was already familiar with the capabilities of an AI tutor due to prior experience participating in similar studies. They had been teaching mathematics at the same school for 16 years at the time of this study.

During their regular math class time (for approximately 20 minutes each day), all students used *Lynnette,* an AI tutor designed for middle school algebra (Figure 1; *center*). In this tutor, students solve equation problems using interactive scaffolding including step-by-step feedback and next-step hints. All students in the study were assigned the same set and sequence of four problems across 12 problem levels (48 problems in total), which started from basic equations and then gradually increased in complexity. Before using the tutor, students worked on a web-based pretest on conceptual and procedural knowledge of algebra. Then, as a post-test, students worked on an isomorphic version of the test after three days of using the tutor. We used two forms, counterbalanced across pre- and post-test so that half the students got the one form as pre-test and the other as post-test, whereas for the other students it was the other way around. The tests contained conceptual items that tested students' conceptual reasoning on problem-solving in algebra [1] and procedural items that asked students to solve algebra problems similar to the ones they practiced in the tutor.

We used Pozyx's UWB (ultrawide-band)-based position sensors to collect the real-time X-Y coordinates of the teacher in the classroom. The positioning system estimates a person's real-time position based on the signal transmitted by UWB tags in a lanyard worn around their neck (Figure 1; *right*) to six anchors (mounted on tripods) along the periphery of the classroom. The positioning system treats the entire classroom as a 2-D coordinate system. Once activated, it samples teacher positions at each second. Next, we measured static coordinates for each student's desk which were mapped to their student IDs in the AI tutor. Any changes to the student seating (a very rare occurrence in this case study) were recorded, so as to allow for accurate tracking of teacher positions in relation to students in the classroom. After time-synchronizing position data with tutor log data for each student, we were able to identify what the student was doing when the teacher was around them. We also measured the static coordinates of all the major objects in the classroom, including the teacher's desk, blackboard, window, and door. These data were used to create visualizations for follow-up interviews with the teacher participating in this study.

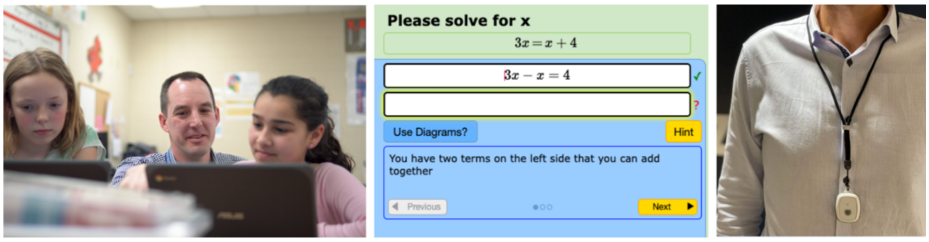

Fig. 1. An illustrative instance of a teacher visit (*left*; image credit: Mathia); An example problem in *Lynnette*, the AI tutor used in the study(*center*); Pozyx tag in a lanyard (*right*; image credit: pozyx.io)

2.2 Teacher Visits in the Temporal Context of Student Learning

Using the teacher's position data and students' seating coordinates, we infer teacher visits to a particular student through a stop detection algorithm. We extended the quantitative definition of *stopping* [15] proposed by adjusting the algorithm parameters to the spatial context of a K-12 classroom (smaller and more densely populated than the open learning spaces from previous studies) and the teaching context of interest (teacher stopping close to a student for individualized attention). The algorithm marks a teacher visit when the teacher's X-Y coordinates are within a small area for a chosen duration of time, *d* or longer. The small area is defined by a circular moving window with the teacher coordinates' centroid as the center and a chosen radius, *r*. Unlike previous studies that set these two parameters (*r* and *d*) ad hoc or using a heuristic, we chose to compute a more accurate combination of parameters with respect to human-coded training data of teacher visits collected in this study. To determine the final set of parameters, we maximized visit recall to 0.17 while constraining the parameter search space during grid search to a precision of at least 0.2, which was more generalizable than unbounded maximization of precision and recall based on cross-validation (see [8] for more details). We also added new logic to detect the student that the teacher is visiting. We define another parameter called range (*rng*). Students with a Euclidean distance smaller than *rng* to the centroid of the detected stop are marked as visited by the teacher. The final set of optimal parameters chosen from this procedure are $d = 21$s, $r = 600$mm (approx. 2 ft), and $rng = 700$mm (approx. 2.3ft). On average, the teacher visited a student once (SD = 2.1) for an average duration of 36 seconds (SD = 77.6s).

Next, we computed the frequency of machine-predicted disengagement measures (i.e., idleness, misuse) and student struggle from AI tutor log data based on models developed in [9]. The models classify the presence of idleness at a threshold of 2 mins, struggle at 25 s (operationalized as slow skill mastery), and misuse at 25 s (operationalized as either hint abuse or rapid sequences of attempts). For each student, each tutor interaction is annotated with indicator variables representing the presence or absence of idleness, tutor misuse, and struggle. These variables can then be aggregated within time windows to compute the percentage of interactions with each behavior. For example, one student may show, on average, idleness at every 10th interaction (0.1) while another student at every 5th (0.2). We similarly aggregate in-system performance measures of learning (i.e., correctness and the average number of errors per problem step) for each student. To investigate the association of teacher visit timing with in-the-moment student needs (i.e., disengagement and struggle), we compare the frequency of these behaviors recorded closer to an upcoming visit (pre-visit) to those closer to a past visit (post-visit). We also created a more granular classification, further splitting up pre- and post-visit interactions. To avoid post-visit interactions overlapping with pre-visit interactions of the next upcoming visit, the more granular classification first assigns each classification a pre- or post-visit class based on whether it is closer to an upcoming or past visit and then classifies its proximity to the closest visit. Specifically, we sample the 50% and 25% closest tutor interactions ahead of teacher visits (closer than 265s and 123s, respectively) and the 50% furthest pre-visit interactions (further away than 265s). We do the same for post-visit interactions with the 25% and 50% closest interactions being classified at 117s and 284s, respectively. We compare these frequencies to a baseline subset of students the teacher never visited (see [3] for code and data).

3 Results

Our analysis was driven by the two research questions presented earlier (see Section 1.2). First (RQ1), we explore factors that may be related to the teacher's choice of students to visit at a given time. Second (RQ2), we investigate the associations between teacher visits and student learning and engagement with the AI tutor.

3.1 RQ1: Factors associated with teacher's choice of students to visit

We hypothesized that the teacher visits are related to both a broader, static measures of student needs such as low prior knowledge (H1) and in-the-moment behavioral indicators from their interaction with the AI tutor (i.e., struggle or disengagement; H2a). We also hypothesized that teachers would additionally prioritize students that struggled or were disengaged the longest relative to other students (H2b).

We find H1 to be not supported by our correlation analysis. Averaging students' scores across procedural and conceptual items on the pre-test, we find that students' prior knowledge was uncorrelated to whether the teacher visited them ($r(48) = -0.22$ [-0.47, 0.06], $p = .118$), the number of visits they received ($r(48) = -0.03$ [-0.31, 0.25], $p = .838$) and the total length of the teacher visits ($r(48) = -0.01$ [-0.28, 0.28], $p = .990$). These findings are based on 50 students instead of 68, as 18 students in the sample were missing test scores. However, we find differences based on the timing of teacher visits concerning students' prior knowledge, as we will elaborate on in our results for H4.

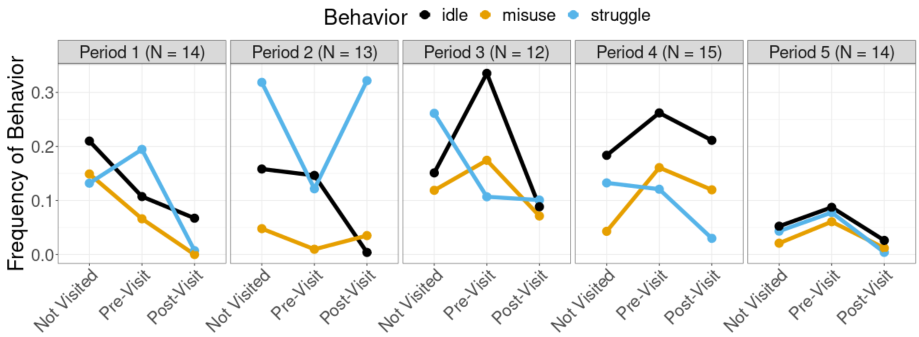

Fig. 2. Frequency of the presence of disengagement and struggle at any given tutor interaction before and after teacher visits per period with students who were never visited as a baseline. Standard error bars are excluded as they are too small to be meaningfully visible.

We find partial support for H2a as there were differential associations with our three behavioral indicators. In the 20-minute sessions, on average, students were idle for 50.5 seconds (SD = 164.3s), struggled for 29.4 seconds (SD = 123.4s), and misused the system for 22.1 seconds (SD = 74.2s). The data are shown in Figure 2, separately for each of the five classes. We compare the struggle and disengagement rates of students the teacher never visited with students' behavior before a visit. The hypothesis would be confirmed if the pre-visit value is higher than the not-visited value. In three out of five classrooms, the teacher tended to allocate visits to students with higher idleness and tutor misuse. By contrast, in three out of five classrooms, the struggle rate was higher in students whom the teacher never visited compared to those who were visited. We note that these comparisons are between individual students, with students that were never visited making up a minority of students (N = 25; 29.07%). Therefore, as an additional way of testing H2a, we conduct additional, one-sided binomial tests. We compare whether students were more likely to show struggle and disengagement right before a teacher visit (recorded as their last tutor interaction before a teacher visit) than the behavioral rates across the whole learning session. We find that, only for idleness, the behavioral frequency was significantly high right before visits ($p = 0.23$, CI95% = [0.17, 0.29] compared to a baseline of $p = 0.14$]. We learned from a follow-up interview with the teacher that different periods have varying characteristics, which might explain the period-level difference shown in Figure 2. According to the teacher, students in Period 1 are not highly-motivated needing them to "stand beside [students] to keep them motivated and working"; Period 3 has most IEP (Individualized Education Program) students that they "quickly figured out the hint strategy and misused the system"; Period 5 is the honors class that is "intrinsically motivated."

Moving on to H2b, we find no indications that the teacher preferentially visited students who exhibited struggle and disengagement earliest in scenarios where multiple students exhibited struggle and disengagement simultaneously. We find that only for 14% of visits, the teacher visited a student who showed the longest (i.e., least recent) period of struggle or disengagement ($p = 0.14$, CI95% = [0.09, 0.20]), including behavioral sequences up to 30 minutes ahead of a visit. There were no significant differences regarding this finding among the three behavioral dimensions across classrooms. Thus, H2b is not confirmed in our data.

3.2 RQ2: Teacher visit associations with student engagement and learning

To test H3, we compare the struggle and disengagement rates of students before teacher visits to after (Figure 2). H3 would be confirmed if the struggle/disengagement is lower post-visit than pre-visit. In all five periods, idleness was significantly lower after a teacher visit. Similarly, tutor misuse, except for Period 2, was significantly lower after a teacher visit. Notably, struggle was lower after teacher visits in all classes except Period 2. Taken together with findings for H2a and H2b, we find that tutor idleness was most robustly associated with the timing of teacher visits (increasing before and decreasing after visits). In line with teachers visits being related to student idlenss, we estimate that per additional SD (i.e., 3.96 mins) of time students spend in the AI tutor without being re-visited by the teacher, the odds of showing idleness in the tutor approximately halves (β = -0.04, OR = 0.55, CI95% = [0.47, 0.64], $p < .001$).

Given heterogeneous associations of our struggle and disengagement measures with teacher visits, we ask which indicators had the strongest association with teacher visits. To answer this question, we calculate the aggregated mean difference between struggle and disengagement metrics before and after teacher visits broken out by period. Given that differences may be larger or smaller depending on how often the behavior occurs generally, we standardize this difference by dividing it by the behavioral frequency in interactions ahead of visits. We find that idleness had the largest standardized change, with idleness being between 37.4% to 97.3% less frequent after than before visits. Averaged by period, this reduction was 59.6%. Struggle had the second largest change; it was reduced by an average of 21.5% (although for period 2, the struggle was increased by 164.0%). Misuse showed an average reduction of 0.6%, ranging from a reduction of 96.3% to an increase of 261.0%. Thus, overall H3 is confirmed for idleness and struggle but not for misuse.

We test associations of teacher visits with student learning in the AI tutor (H4). We do this by associating teacher visits with in-system performance and learning gain on out-of-tutor tests. We begin by reporting relations between teacher visits and the two measures of in-system performance, correctness of individual student responses in the tutor and the average number of attempts per step. We compare logistic regression models via likelihood-ratio tests to investigate whether the correctness of students' first attempt at each step is significantly different pre- and post-visits (as operationalized via our time binning described in Section 2.2). We compare a model featuring an indicator variable representing whether the given attempt is the student's first attempt at the step with a model additionally featuring teacher visit timing represented through the time binning variable. We find that adding teacher visit timing significantly improved model fit ($\chi2(6) = 41.58$, $p < .001$). In addition, adding prior knowledge (measured as pre-test score) as an additive effect ($\chi2(1) = 541.21$, $p < .001$) and in interaction with teacher visit timing ($\chi2(6) = 11.29$, $p = .080$) further improved model fit. We follow a similar procedure using Poisson count models inferring the number of attempts at a given step. Here, a model with just an intercept served as a baseline. Adding teacher visit timing ($\chi2(1) = 2541.90$, $p < .001$), prior knowledge ($\chi2(1) = 5781.50$, $p < .001$), and the interaction of both ($\chi 2(1) = 263.20$, $p < .001$), significantly improved model fit. Like before, we omitted 18 cases due to missing test scores in this part of the analysis. Taken together, these findings highlight the utility of the relative timing of teacher visits in student in-tutor correctness, although the association was not uniformly positive, as assumed by H4, visualized in Figure 3.

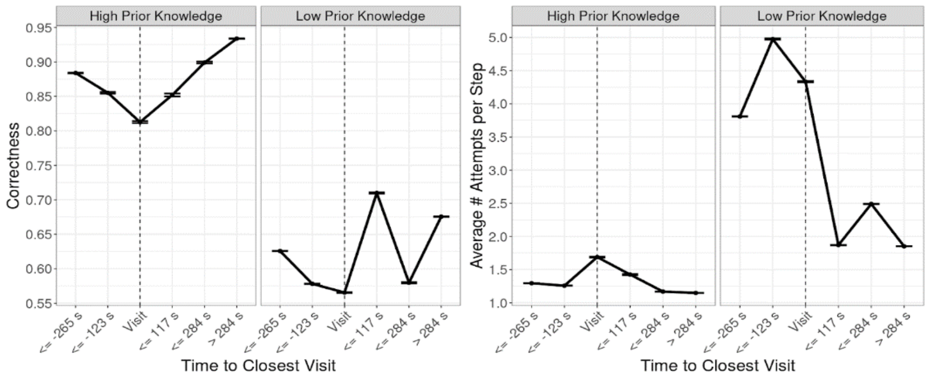

Fig. 3. Average correctness and the number of attempts per problem step broken out by the relative timing of tutor interactions with respect to teacher visits and whether students exhibited an above-median prior knowledge in the pre-test including 2 SE bars. Observations to the right of the dashed line designate aggregations of data points closer to past than upcoming visits.

Based on Figure 3, we find two notable patterns. First, while all visited students exhibited decreasing correctness before teacher visits, students with high prior knowledge continuously increased their correctness at steps after a teacher visit. On the other hand, low prior knowledge students only had correctness improvements after teacher visits for around two minutes before falling back in correctness, except after around five minutes if the teacher did not revisit them. Second, high prior knowledge students only exhibited an increased number of attempts per step two minutes ahead of a teacher visit. In comparison, students with low prior knowledge already exhibited that behavior more than four minutes before a visit. In other words, students with high prior knowledge were visited by the teacher earlier after their number of attempts per step started to rise.

Finally, to continue our test of H4, we investigate the association of teacher visits, that is, their number and length, with learning as measured by pre/post tests. Including the length of visit in this analysis was motivated by observing considerable variance in the duration of teacher visits ($M = 36.85$ s, $SD = 18.53$ s). Given heterogeneous associations between teacher visits and in-tutor performance as displayed in Figure 3, we e select a set of control variables through an *AIC*-based backward search for Gaussian regression models that predict pre/post procedural and conceptual learning gains, to isolate the effects of the number and length of visits. Positive marginal effects of the number and length of visits given these control variables would confirm H4. The procedure removes the least significant feature from a model including all features until there is no further improvement in *AIC*. Our feature search space includes features representing student struggle and disengagement (e.g., the average number and length of idleness sequences), student prior knowledge, and whether students were missing on the three days of data collection. We find that the number of teacher visits was significantly negatively associated with procedural (β = -0.31, CI95% = [-0.56, -0.06], p = .018) and conceptual (β = -0.95, CI95% = [-1.78, -0.12], p = .045) learning gain (see Section 2.1 for definitions of these measures). In addition, the total time spent with a student was significantly positively associated with conceptual learning gain (β = 1.02, CI95% = [0.06, 1.99], p = .038). Overall, this analysis reveals that teacher visits correlate with student leaning gains, though not in a straightforward manner. We unpack these estimated effects. First, increasing the number of visits while keeping the total length of visits constant (i.e., students experiencing more, shorter visits) was associated with lower procedural and conceptual learning gain. Second, increasing the total length of visits while keeping the number of visits constant (i.e., students experiencing fewer, longer visits) was associated with higher conceptual learning gains but not higher procedural learning gains. In a follow-up interview, the teacher noted that the duration of their visits varied and that some students needed short but frequent attention for reassurance, while others needed longer assistance on conceptual understanding. Taken together, while teacher visits exhibited significant associations with student learning, as measured by in-tutor performance and out-of-tutor knowledge tests, our hypothesis that this association would be uniformly positive was not supported.

4 Discussion and Conclusion

In recent years, the AIED community has increasingly focused on developing teacher tools and dashboards for AI-enabled classrooms. As we progress in such work on augmenting teachers' practices around AI tutors, it could be highly beneficial to have a deeper theoretical and empirical understanding of decisions teachers make in the physical classroom and their impact on student learning and experience with AI tutors. In this paper, we investigate spatiotemporal aspects of student-teacher interactions to better understand how teachers support student learning with AI tutors. Our analysis reveals that teacher visits were associated more with students' in-the-moment behavioral indicators (e.g., idleness) than with a broader, static measure of student needs such as low prior knowledge. While teacher visits were often associated with positive changes in student behavior afterward (e.g., decreased idleness), it was not clear that the teacher always selected a student who may have needed a visit the most (e.g., sometimes there were students who had been disengaged for much longer than the visited student; students with high prior knowledge were often visited earlier than low prior knowledge students). Furthermore, we find that *longer* teacher visits were associated with higher conceptual learning gain, while *frequent* visits were associated with lower conceptual and procedural learning gain. The latter might represent a selection effect, not an indication that the teacher's help was detrimental. Students who have difficulty with the material may attract many teacher visits. Therefore, there may be qualitative differences between teacher visits that may relate to conceptual and procedural support that calls for further scrutiny in future research. Lastly, we also observe differences in after-visit patterns between different student groups that potentially are not desirable. Understanding student-level differences in the antecedents and impact of teacher visits remains an important topic for future research.

The empirical contributions of this study have several implications for the development of teacher support for orchestration and reflection. First, the insights derived suggest that it may be useful if future teacher support tools were to present to teachers information about how their spatial pedagogical practices in the classroom (and visits, specifically) relate to student learning and engagement. Second, our findings also suggest that teacher behavior may not be ideal in some cases (e.g., teachers may have missed visiting struggling students), further motivating the need for well-designed tools to improve teachers' awareness and sensemaking [9]. Third, our study provides some new forms of analytics that, embedded in teacher tools, could help teachers better prioritize their limited classroom time between conflicting student needs (e.g., analytics that help identify students who were disengaged for longer). Fourth, the quantitative conceptualization of teacher practices opens up new possibilities for teaching analytics focused on helping teachers reflect on their own practices to make pedagogical improvements. As a preliminary illustration of that idea, while reflecting on a visualization of their visits, the teacher from the current case study said in one of the follow-up interviews, "Look at [student's name], I barely stopped by him, and he's the kid who is struggling, but I was hardly there. So, I want to check in on him a little bit more often." They were referring to a student who was seated at the back of the classroom. Such reflections led the teacher to imagine concrete classroom enactments, such as potential changes to the seating arrangement to have better visibility to students who may misuse the system, to have a peer tutor in close proximity to struggling students, and to group multiple struggling students to provide conceptual intervention.

In addition, there are several methodological implications of the current study for AIED research. First, combining data sources from the physical and virtual spaces enables an integrated, temporal analysis of student behaviors in the AI tutor and student-teacher behaviors in the classroom. Such analyses are necessary if we aim to improve our understanding of teachers' roles and practices in supporting students' learning with AI tutors. Second, the current study demonstrates the feasibility of collecting and analyzing such data in an automated way. To the best of our knowledge, the current study is the first to collect fine-grained, time-synchronized data on teacher position and log data of student-tutor interaction in an AI-enabled classroom in an automated way. Unlike other automated approaches like videos used in previous studies to quantify student-teacher interactions [17], position sensing is less intrusive, preserves student privacy, has a lower risk of unintended surveillance [4], and requires lower post-processing [18] to generate position coordinates. Third, we demonstrate how an existing algorithm (stop detection; [8]) could be extended to quantitatively conceptualize a teaching construct of interest (visit) in a K12 classroom.

The study has several limitations that point to interesting opportunities for future work. First, our analysis revealed differences in the nature of visits (e.g., infrequent long visits versus frequent short visits). While we explored visit length in this study, future research needs to further contextualize teacher visits, for example, by encoding their content (e.g., help-giving, socioemotional support, reassurance). Second, the quantitative definition of teacher visits in this paper does not capture student-teacher interactions happening without physical proximity and may misattribute instances when a teacher stands near one student but talks to another, or observes the class, or does nothing. Third, our exploratory analysis doesn't consider other factors in an ecological setting like a classroom that may play a role (e.g., students seeking help by raising their hand, students getting help from a peer). Lastly, this is a case study with one teacher and limited time. The findings need to be tested broadly for generalizability.

Our findings confirm that teachers play an important role in AI-enabled classrooms, but also that they are limited in their abilities to perceive and prioritize *all* student needs in real-time. We have seen that tools designed to improve teachers' awareness and sensemaking of students' learning with AI tutors lead to better student learning [9]. Our results inform future design of teacher support tools that intentionally bring out the best of teacher abilities and overcome some of their limitations.

References

1. Crooks, N. M., Alibali, M. W.: Defining and measuring conceptual knowledge in mathematics. Developmental review, 34(4), 344-377 (2014).
2. Schofield, J. W., Eurich-Fulcer, R., & Britt, C. L.: Teachers, computer tutors, and teaching: The artificially intelligent tutor as an agent for classroom change. American Educational Research Journal, 31(3), 579–607 (1994).
3. https://pslcdatashop.web.cmu.edu/Project?id=879, last accessed 2023/05/07.
4. Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Sherin, B. L.: Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. The journal of the learning sciences, 19(1), 3-53 (2010).
5. Fernandes, A. C., Huang, J., Rinaldo, V.: Does where a student sits really matter?-The impact of seating locations on student classroom learning. IJAES 10, 1 (2011).
6. Stang, J. B., Roll, I.: Interactions between teaching assistants and students boost engagement in physics labs. Physical Review Special Topics-Physics Education Research, 10(2) (2014).
7. Dessus, P., Cosnefroy, O., & Luengo, V.: "Keep your eyes on'em all!": A Mobile eye-tracking analysis of teachers' sensitivity to students. EC-TEL*,* 72-84 (2016).
8. Shou, T., Borchers, C., Karumbaiah., S., Aleven, V: Optimizing Parameters for Accurate Position Data Mining in Diverse Classrooms Layouts. EDM (2023).
9. Holstein, K., McLaren, B. M., Aleven, V.: Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In: AIED. pp. 154–68. Springer (2018).
10. Holstein, K., McLaren, B. M., Aleven, V.: SPACLE: Investigating learning across virtual and physical spaces using spatial replays. In: Proceedings of the LAK, 358-67 (2017).
11. Kessler, A., Boston, M, Stein, M. K.: Exploring how teachers support students' mathematical learning in computerdirected learning environments. ILS 121(1/2), 52–78 (2019).
12. Knight, S., Wise, A. F., Chen, B.: Time for change: Why learning analytics needs temporal analysis. Journal of Learning Analytics, 4(3), 7-17 (2017).
13. Knoop-van Campen C. A. N., Wise, A., Molenaar, I.: The equalizing effect of teacher dashboards on feedback in K-12 classrooms. Interact. Learn. Env. 1–17 (2021).
14. Lim, F. V., O'Halloran, K. L., Podlasov, A.: Spatial pedagogy: Mapping meanings in the use of classroom space. Cambridge Journal of education 42, 2, 235–251 (2012).
15. Martinez-Maldonado, R., Echeverria, V., Mangaroska, K., Shibani, A., Fernandez-Nieto, G., Schulte, J., Shum, S. B.: Moodoo the tracker: Spatial analytics for characterising classroom pedagogies (2020).
16. Mei, C. C. Y., Chin, H. B., Taib, F.: Instructional proxemics and its impact on classroom teaching and learning. IJMAL 1, 1, 69–85 (2017).
17. van Es, E. A., & Sherin, M. G.: Expanding on prior conceptualizations of teacher noticing. ZDM – Mathematics Education, 53(1), 17–27 (2021).
18. Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., Dillenbourg, P.: Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. Journal of computer assisted learning, 34(2), 193-203 (2018).