

# Who to Help? A Time-Slice Analysis of K-12 Teachers' Decisions in Classes with AI-Supported Tutoring

Qiao Jin  
Carnegie Mellon University  
Pittsburgh, PA, USA  
[georgiej@andrew.cmu.edu](mailto:georgiej@andrew.cmu.edu)

Stephen Fancsali  
Carnegie Learning, Inc.  
Pittsburgh, PA, USA  
[sfancsali@carnegielearning.com](mailto:sfancsali@carnegielearning.com)

Conrad Borchers  
Carnegie Mellon University  
Pittsburgh, PA, USA  
[cborcher@cs.cmu.edu](mailto:cborcher@cs.cmu.edu)

Vincent Aleven  
Carnegie Mellon University  
Pittsburgh, PA, USA  
[aleven@cs.cmu.edu](mailto:aleven@cs.cmu.edu)

## ABSTRACT

Classroom orchestration tools allow teachers to identify student needs and provide timely support. These tools provide real-time learning analytics, but teachers must decide how to respond under time constraints and competing demands. This study examines the relationship between indicators of student states (e.g., idle, struggle, system misuse) and teachers' decisions about whom to help, using data from 15 classrooms over an entire school year (including 1.6 million student actions). We explored (1) how student states relate to teacher intervention, especially when multiple students need help, (2) how learning rates and initial proficiency affect the likelihood of receiving help, and (3) whether teachers prioritize student states that align with system help-seeking patterns. Using a time slice analysis, we found that teachers primarily helped students based on idleness, while student help seeking in the tutoring system was primarily related to struggle. Furthermore, our findings show that students' receipt of help from teachers was significantly positively correlated with their in-system learning rate. These findings highlight how learning analytics of student states can enhance teacher support in AI-supported classrooms and assess the effectiveness of teacher support, offering insights into key indicators for orchestration tools.

## Keywords

Classroom orchestration, intelligent tutoring systems, teacher help, detectors, help-seeking, struggle

## 1. INTRODUCTION

Classroom teaching is a complex task that requires educators to continuously observe each student's learning state to determine who needs support and what type of assistance would be most effective [19, 3, 9]. Teachers must align their instructional actions with students' needs to achieve their

Qiao Jin, Conrad Borchers, Stephen Fancsali, and Vincent Aleven. Who to Help? A Time-Slice Analysis of K-12 Teachers' Decisions in Classes with AI-Supported Tutoring. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette (eds.) Proceedings of the 18th International Conference on Educational Data Mining, Palermo, Italy, July, 2025, pp. 640–645. International Educational Data Mining Society (2025).

© 2025 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

<https://doi.org/10.5281/zenodo.1587029>

pedagogical goals [16]. With the advancement of AI, AI-driven classroom orchestration tools have been developed to enhance teachers' instructional practices during live classroom activities, particularly by directing their attention to students who need help the most (e.g., Lumilo [14], MT-Dashboard [24], and DREAM [28]). Research has shown that when AI recommendations are used to prioritize learners in need, teachers can adjust their instructional strategies accordingly to focus on those who need the most support, leading to measurable improvements in student learning and performance [14, 2, 10]. Additionally, AI-powered learning analytics can offer a broader view of classroom dynamics, enabling educators to personalize instruction and manage learning activities more efficiently [3, 24, 12].

Despite the rise of real-time teaching orchestration tools and learning analytics systems that help teachers better understand student learning, most research remains limited to small-scale studies and research-controlled classrooms [9]. A prior study [9] has investigated software detected student states (idle, monitor, working, and offline) when receiving help but did not model their overall impact on teacher intervention practices. By modeling between student information (e.g., learning states) and teacher decision, we can gain insight into teacher decision-making patterns and investigate which student conditions influence teachers' choices of whom to visit. By understanding what guides teacher behavior, we can identify what analytics could pose additional opportunities to guide teacher prioritization of what student to help.

In this study, we examine the relationship between teacher decisions about whom to help in classes and student states using an AI-powered classroom orchestration tool, LiveLab [11]. LiveLab has been widely adopted by educators and provides real-time insights into student progress within MATHia, a mastery-based intelligent tutoring system used by over 600,000 students in grades 6–12 across the United States for mathematics education. Given that students progress through the curriculum at different paces, teachers often struggle to monitor individual learning trajectories and identify students in need of assistance. To support student learning, MATHia allows students to request hints from the system. Additionally, LiveLab offers a dashboard that shows students recent

progress and activity in MATHia (e.g., idle, working), also enables teachers to manually log instances of providing help to students within its detailed analytics interface. We employ rule-based detectors to classify when students exhibit four key engagement and performance states: doing well, idle, struggling (sometimes referred to as wheelspinning [6]), and system misuse (sometimes referred to as gaming the system [4]). We chose these four indicators because past co-design research on classroom orchestration tools with K-12 teachers revealed that teachers favor these four indicators among other student detectors and find them most informative for which student to help [15]. Our research explores how educators integrate this AI-driven tool into their daily teaching practices, with a specific focus on the “marked helped” feature, where teachers support students individually for brief moments of time while they learn with technology (and indicate that they have done so in LiveLab’s dashboard). We focus on these help moments because they are known to play an important role in student learning and the effectiveness of tutoring systems [18, 26].

By analyzing teacher-logged decisions of whom to help across a diverse range of classroom settings—covering 15 unique classes in which this teacher logging of support instances was used most frequently, throughout the academic year and more than 1.6 million student-system interactions—we gain valuable insights into how teachers engage with AI-detected student learning and engagement states in real-world environments. Specifically, our study investigated the following research questions:

**RQ1:** How do student states relate to teacher intervention in general (RQ1a) and during instructionally challenging moments (where more than one student needs help) (RQ1b)?

**RQ2:** How do students learning rate and initial proficiency relate to the likelihood of receiving teacher help?

**RQ3:** Do the student states that teachers prioritize for intervention align with those most correlated with student help-seeking in the tutoring system?

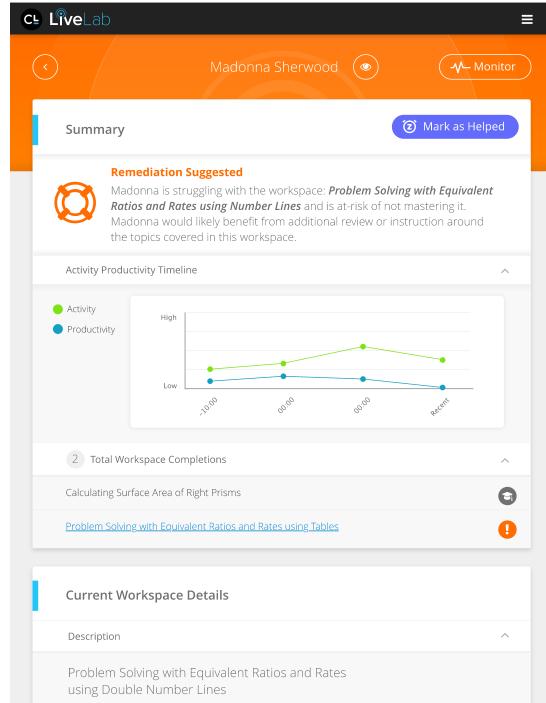
## 2. METHOD

### 2.1 Dataset

Student interaction data was collected from MATHia [11], comprising 1,641,887 transactions (i.e., student actions) from 358 students enrolled in 15 middle and high school math classes across ten U.S. K-12 schools during the 2022-23 school year. These classes were selected based on the relative frequency with which teachers used a feature to mark moments of help, meaning the teacher reported through the LiveLab dashboard<sup>1</sup> (Figure 1) when they helped a specific student after 1:1 help was provided. These “marked as helped” instances were then recorded in the log data with timestamps, together with the student ID, so we can relate them to students’ actions also recorded in the log data.

### 2.2 Detectors and Feature Engineering

<sup>1</sup>MATHia is typically part of a schools’ core math curriculum and instruction in a blended model that recommends usage MATHia for 40% of instructional time during a typical week of math instruction (e.g., two of out five instructional math periods). LiveLab is provided as optional support for orchestrating this MATHia classroom time.



**Figure 1:** LiveLab dashboard shows real-time student progress monitoring and remediation suggestions for teacher. Teacher could manually log instances of providing help to students using a button “Mark as Helped”.

As is common practice for tutoring systems, the log data included timestamped interactions of students with the system, including graded attempts and hints requests. These log data were enhanced with behavioral detectors to identify students’ learning states, using LearnSphere [27], a collaborative open data infrastructure that supports the sharing, analysis, and modeling of large-scale educational data for learning science research. Behavioral detectors were recoded into binary variables, indicating the presence or absence of each student state.

We classify students’ learning states using rule-based detectors, identifying four key categories: idle (no interactions in the software for 2 minutes or more [14]), system misuse (rapid guessing, abusing hints, or gaming the system [4]), struggle (students attempting to master a skill three or more times without mastering it despite system support [6]), and doing well (low recent error rate in current activity, defined as an error rate less than 20% over the last 10 actions [14]). These detectors were based on open-source code featured in past research on teacher augmentation tools, which have shown to significantly improve student learning gain when shown to teachers in classrooms where students learn with tutoring systems [14, 15]. These detectors capture different aspects of student learning behaviors, which are crucial for understanding how teachers respond to students [11, 18].

To further capture student learning dynamics, we considered two additional factors: (1) System hint usage: This controls student help-seeking behavior within the system and may indicate a need for instructional support. We define it as

whether a student requests system-provided hints, serving as a baseline measure for identifying students who require additional instruction beyond system feedback [1]. (2) Challenging moments: It is possible that teacher help behavior not only depends on whether or not a student needs help, but also how many other students need help. For example, a teacher may have to make a decision between multiple students needing help at the same time. We define these situations as instructionally “challenging moments” – whether more than one student exhibits disengagement or struggle within 15-minute intervals (i.e., idle, system misuse, or struggle). The time interval follows the next section’s model, tested for robustness across different time slices.

### 2.3 Statistical Modeling

When estimating the relationship between students’ states and help-seeking with teacher help, there is a natural variation in when and how teachers report which student they have helped. In discussion with the data provider and past classroom research [18], we decided to analyze our data in time slices, meaning we estimate the co-occurrence of teacher behavior and student behavior within timeframes within the class periods, typically lasting about 45 minutes. Similar co-occurrence or time slicing analyses have been applied to study teacher behavior in tutoring system classrooms [17] and also have been used to understand diagnostic behavior in other domains [7].

Based on these considerations, a 15 minute time slice window was chosen and all student binary detectors and teacher help decisions were aggregated within these windows, with 1 coded as a given behavior being present at any point of the time slice and 0 otherwise. We checked robustness of results for different time slice intervals ranging from 5-20 minutes and report results for 15 minute windows as they offer a good balance between data sparsity and statistical power (as larger windows lead to sparser and potentially noisier data with fewer occurrences, but also more statistical power through more observations [7]).

To assess the relationship between student behaviors and receiving help (RQ1a), a generalized linear mixed effects model (GLMM) was estimated using logistic regression [21]. The model included random intercepts for students nested in classes to account for hierarchical dependencies and student-differences in help receiving (teacher intervention) given repeated measurements introduced through time slicing. The final model evaluated the likelihood of receiving help based on whether help was requested in the system (i.e., hint usage) and a set of detected behaviors:

$$\begin{aligned} \text{TeacherIntervention} \sim & (1|\text{Student}) + (1|\text{Class}) \\ & + \text{HintUse} + \text{Idle} + \text{Misuse} \quad (1) \\ & + \text{Struggle} + \text{DoingWell} \end{aligned}$$

The model was assessed using intraclass correlation coefficients (ICCs) to examine variance explained at the student and class levels as well as standard Wald-test of model main effects.

To investigate student states related to teacher intervention during challenging moments, we examined the interaction effects of challenging moments (binary indicator) with all

main effects of Equation 1 by estimating a *separate* model to answer RQ1b.

To answer RQ3 that whether student state that most indicative of students requiring teacher intervention align with the student help-seeking in the tutoring system, we re-used our model 1 but modeling student help-seeking with tutoring system as the binary outcome instead of a predictor.

### 2.4 Individual Differences in Learning

Past research has collected evidence that teacher help improves learning with tutoring systems [18, 14]. To extend this evidence and answer RQ2, we studied how teacher support was correlated with student learning rates (as a measure of learning efficiency) and estimated prior knowledge. To explore student-level differences in initial proficiency, an iAFM learning model [22] was estimated to track student progress over time on each skill using the standard skill model [8]. The standard iAFM model is in Equation 2. The model infers the correctness of the first attempts at a problem-solving step ( $Y_{correct}$ ) based on an individualized mixed-model student intercept representing students’ initial proficiency ( $\tau_{stud.}$ ), an intercept per skill representing initial skill difficulty ( $\tau_{skill}$ ), and an opportunity count slope ( $\beta_{opportunity}$ ) representing overall learning rate [20]. Further individualized student slopes model student learning rate differences ( $\tau_{stud./opportunity}$ ).

$$Y_{correct} = \tau_{stud.} + \tau_{skill} + \beta_{opportunity} + \tau_{stud./opportunity} + \epsilon \quad (2)$$

Random intercepts (representing initial student proficiency) and learning rates for students were extracted and correlated with student-specific effects from the help-receiving model to examine potential associations between student learning attributes and variation in student help-receiving behaviors. All statistical analyzes were performed in R using the lme4 package [5]. Our code is publicly available on GitHub<sup>2</sup>.

## 3. RESULTS

### 3.1 Factors Associated with Teacher Intervention (RQ1)

#### 3.1.1 Teacher Intervention in General

The logistic regression model assessing predictors of receiving teacher help indicated that students who requested help in the system through hints had significantly higher odds of receiving teacher assistance ( $OR = 1.70$ , 95% CI [1.37, 2.11],  $p < .001$ ). Additionally, students who were idle were significantly more likely to receive help compared to those in other states ( $OR = 3.26$ , 95% CI [1.42, 7.50],  $p = .005$ ). However, system misuse ( $OR = 0.79$ , 95% CI [0.28, 2.23],  $p = .654$ ), struggle ( $OR = 0.78$ , 95% CI [0.22, 2.75],  $p = .704$ ), and students doing well ( $OR = 0.81$ , 95% CI [0.32, 2.02],  $p = .650$ ) were not significant predictors of receiving help.

Further decomposition of variance at different levels indicated that 16.9% of the variance in receiving help was attributable to differences between students (ICC = 0.169), while 6.8% of the variance was explained by class-level differences (ICC = 0.068). The univariate odds ratios for the individual predictors were generally consistent with the magni-

<sup>2</sup><https://github.com/GeorgieQiaoJin/A-Time-Slice-Analysis-of-Teachers-Decisions.git>

tude and direction of the effects observed in the main model. This consistency reinforces the robustness of the model in capturing the factors associated with teacher help-selection decisions.

### 3.1.2 Teacher Intervention in Challenging Moments

Results show that teachers were more likely to provide help during challenging moments ( $OR = 2.26$ , 95% CI [1.09, 4.67],  $p = .028$ ). Additionally, we analyzed the interaction effects between challenging moments and detected uses. Specifically, the interactions between challenging moments and system misuse ( $OR = 1.63$ , 95% CI [0.18, 14.40],  $p = .659$ ), struggle ( $OR = 1.59$ , 95% CI [0.11, 22.14],  $p = .730$ ), and doing well ( $OR = 2.46$ , 95% CI [0.36, 16.98],  $p = .362$ ) were not statistically significant. However, there was a marginal association whereby teachers were less likely to help students based on their idleness ( $OR = 0.14$ , 95% CI [0.02, 1.05],  $p = .056$ ) when challenging moments occurred.

## 3.2 Individual Differences in Learning (RQ2)

Given the relatively large ICC for students, we examined whether individual differences characteristics were associated with variations in help reception, focusing on students' initial proficiency and learning rate. No significant correlation was found between students' initial proficiency and their intercept from the help-receiving model ( $r = -0.067$ , 95% CI [-0.174, 0.042],  $p = .226$ ), suggesting that students' baseline skill levels were not systematically associated with their probability of receiving help. However, a significant positive correlation was observed between the intercept of the help-receiving model and students' learning rate ( $r = 0.125$ , 95% CI [0.016, 0.230],  $p = .024$ ). This suggests that students who had a higher propensity to be visited by teachers tended to learn more from the instruction of the tutoring system.

## 3.3 Alignment of Teacher Priorities and Student Help-Seeking (RQ3)

The results of our logistic regression model on system help-seeking behavior indicate that students who were struggling had significantly higher odds of help-seeking in the tutoring system ( $OR = 2.46$ , 95% CI [1.37, 4.43],  $p = .003$ ). Conversely, idleness ( $OR = 0.93$ , 95% CI [0.62, 1.38],  $p = .703$ ) and system misuse ( $OR = 1.11$ , 95% CI [0.75, 1.64],  $p = .596$ ) were not significant predictors of help-seeking in the system. Similarly, doing well showed no significant difference in help-seeking behavior ( $OR = 1.11$ , 95% CI [0.79, 1.56],  $p = .553$ ). These suggest that students who were struggling are more likely to seek help in the system, while idle behavior and system misuse do not strongly predict system help-seeking. In other words, there was an empirical mismatch between instructional needs of students and teacher 1:1 support, which was more guided by student disengagement (idleness).

## 4. DISCUSSION

Understanding how teachers allocate their limited time and attention is critical for optimizing classroom orchestration tools [14, 18]. This study provides novel insights into teacher intervention patterns in AI-supported classrooms by analyzing over 1.6 million student-system interactions across 15 classrooms in K-12 classrooms. We examined the role of AI-generated student engagement indicators in predicting

teacher help and system hint usage, offering an empirical assessment of how teachers respond to student disengagement, help-seeking, and struggle and whether those align with student needs.

Our results reveal several factors associated with teacher intervention to answer RQ1: in general, when students used more system hints and exhibited idleness were significantly more likely to receive teacher assistance (RQ1a). This finding is consistent with previous research that analyzed middle school classrooms using MathTutor, where teachers' helping decisions were most associated with student idleness [18]. However, system misuse, struggle, and student doing well were not significant predictors of teacher help. Based on past observation studies [18, 15], it is likely that teachers cannot easily tell whether a student is struggling or misusing the system due to classroom management demands. As a consequence, teacher may pay more attention to openly-visible engagement issues rather than cognitive needs of students. Rather than responding to cognitive needs, teachers are known to "make rounds" to check in and resolve content issues [26]. In contrast, targeted help of students during struggle, as struggle, which was most associated with student help-seeking in the system, could help allocate teacher help more effectively [23]. Therefore, our findings provide novel evidence that analytics guiding teachers to struggling students could unlock additional opportunities to improve classroom practice with tutoring systems. Notably, although some detectors (e.g., system misuse and struggle) did not reach statistical significance, this does not mean that these factors are unimportant for optimal intervention strategies. Prior research suggests these factors (e.g., gaming behavior [13] (akin to misuse in our study), frustration (akin to struggle in our study) [25]) still significantly correlated with student learning outcomes when learning with software.

During challenging moments (RQ1b), teachers were more likely to provide help. This aligns with teaching logic that, when more students require help, teachers tend to intervene more proactively. However, the weaker predictive power of idleness in these situations suggests that teachers may rely more on other signals to decide whether to intervene. For example, teachers more generally visited more students when multiple students required help. It is possible that they attempted to quickly visit several students to manage their classroom efficiently [26], however, that hypothesis is subject to future research. Future directions can explore how teachers intervene for different detected behaviors. For example, prior studies have shown that idle students primarily received encouragement and some technical support [9] – often just enough to quickly get them back on task. This might explain why idle students were more likely to receive help, as the time cost of assisting them is lower.

Prior work found that teacher visits helped low-learning-rate students adopt more effective problem-solving behaviors [8]. Our results to answer RQ2 align with this, showing that students with higher learning rates were more likely to receive teacher help, while initial proficiency was not a predictor. This suggests that teacher attention is influenced by students' responsiveness to instruction rather than just baseline ability. Future work should explore whether early conceptual support for low-learning-rate students can fur-

ther enhance learning outcomes.

Our findings reveal that student states most correlated with help-seeking, and additional instructional need during cognitive tutoring may not always align with the states teachers pay attention to when deciding whom to provide with one-on-one teacher support (RQ3). Specifically, while struggle was most associated with student hint seeking in the system, idleness was most associated with teacher help-selection decisions. This suggests that teachers may underestimate or overlook the need for assistance among struggling students, as these students tend to (unsuccessfully) rely on system hints rather than waiting for teacher intervention. Therefore, when designing AI-driven classroom suggestion systems, it may be necessary to adjust the weighting of different behavior detectors to better meet students' needs. ITS systems could also integrate these findings to optimize help-triggering mechanisms. For example, detect when students are idle but delay notifying teachers until a more appropriate moment.

#### 4.1 Limitations and Future Work

We lack action data on offline students (although they can still be marked as helped in the system), preventing us from linking teachers' helping decisions to their offline actions. This missing data is unlikely to bias the results as students who were offline more often were also likely to be more disengaged. If anything, our analysis may underestimate the effect of disengagement by missing instances of struggle or idleness, but were not logged into the system. Future work could incorporate qualitative teacher observations or alternative engagement metrics to better capture offline behavior. Another limitation is that teachers had access to the "idle" status in the dashboard, which may have influenced their decision making. This visibility could mean that teachers either actively responded to idleness in real time or relied on system detections over their own observations. As a result, recorded decisions might partially reflect system guidance rather than independent teacher recognition. Future studies should examine how system-provided indicators shape teacher behavior and whether helping decisions change when automated engagement signals are unavailable. Furthermore, we selected only classes where teachers logged support instances most frequently, but, despite this filtering, the completeness of these logs is unclear. Some real-time decisions may have gone unrecorded, which could limit the generalizability of our findings to settings with different logging practices. Future work may consider defining inclusion criteria based on completeness checks (e.g., through classroom observations or teacher self-reports). In addition, the dataset does not include any further information about the instructors. Future work should consider whether instructional contexts, levels of experience and preparation make teachers more or less likely to use LiveLab or to use it in their instructional practice in different ways.

### 5. CONCLUSION

This study investigated how student states and system help-seeking behaviors relate to teacher intervention in K-12 classrooms. Using a large-scale dataset of over 1.6 million student interactions, we found that teachers were significantly more likely to assist students exhibiting idleness than those struggling academically, even though struggle was more strongly

associated with student help-seeking in the tutoring system. This mismatch suggests that teachers may respond more to visible disengagement than to deeper cognitive challenges. Moreover, students with higher learning rates were more likely to receive help, suggesting that teacher attention may be driven more by perceived responsiveness than by baseline proficiency. Overall, this work contributes to understanding current teacher intervention practices and to providing design and research implications for future teacher awareness and recommendation systems aimed at improving classroom support strategies.

### 6. ACKNOWLEDGMENTS

This research was funded by the Institute of Education Sciences (IES) of the U.S. Department of Education (Award #R305A240281). We also thank Carnegie Learning, Inc. for providing access to the data used in this study.

### 7. REFERENCES

- [1] V. Aleven, I. Roll, B. M. McLaren, and K. R. Koedinger. Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 26:205–223, 2016.
- [2] P. An, S. Bakker, S. Ordanovski, R. Taconis, C. L. Paffen, and B. Eggen. Unobtrusively enhancing reflection-in-action of teachers through spatially distributed ambient information. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–14, 2019.
- [3] P. An, K. Holstein, B. d'Anjou, B. Eggen, and S. Bakker. 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*, pages 1–17, 2020.
- [4] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner. Off-task behavior in the cognitive tutor classroom: When students' game the system". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 383–390, 2004.
- [5] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- [6] J. E. Beck and Y. Gong. Wheel-spinning: Students who fail to master a skill. In *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16*, pages 431–440. Springer, 2013.
- [7] C. Borchers, T. F. Eder, J. Richter, C. Keutel, F. Huettig, and K. Scheiter. A time slice analysis of dentistry students' visual search strategies and pupil dilation during diagnosing radiographs. *PLoS One*, 18(6):e0283376, 2023.
- [8] C. Borchers, Y. Wang, S. Karumbaiah, M. Ashiq, D. W. Shaffer, and V. Aleven. Revealing networks: understanding effective teacher practices in ai-supported classrooms using transmodal ordered network analysis. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pages 371–381, 2024.
- [9] E. Brunskill, K. A. Norberg, S. Fancsali, and S. Ritter. Examining the use of an ai-powered teacher

- orchestration tool at scale. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pages 356–360, 2024.
- [10] S. E. Fancsali, K. Holstein, M. Sandbothe, S. Ritter, B. M. McLaren, and V. Aleven. Towards practical detection of unproductive struggle. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21*, pages 92–97. Springer, 2020.
  - [11] S. E. Fancsali, M. Sandbothe, and S. Ritter. Orchestrating classrooms and tutoring with carnegie learning's mathia and livelab. In *Human-AI Math Tutoring@ AIED*, pages 1–11, 2023.
  - [12] S. Feng, L. Zhang, S. Wang, and Z. Cai. Effectiveness of the functions of classroom orchestration systems: A systematic review and meta-analysis. *Computers & Education*, 203:104864, 2023.
  - [13] Y. Gong, J. Beck, N. T. Heffernan, and E. Forbes-Summers. The impact of gaming (?) on learning at the fine-grained level. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS2010) Part*, volume 1, pages 194–203, 2010.
  - [14] K. Holstein, B. M. McLaren, and V. Aleven. Student learning benefits of a mixed-reality teacher awareness tool in ai-enhanced classrooms. In *Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part I 19*, pages 154–168. Springer, 2018.
  - [15] K. Holstein, B. M. McLaren, and V. Aleven. Co-designing a real-time classroom orchestration tool to support teacher–ai complementarity. *Journal of Learning Analytics*, 6(2):27–52, 2019.
  - [16] A. R. James. *Elementary physical education teachers' and students' perceptions of instructional alignment*. University of Massachusetts Amherst, 2003.
  - [17] S. Karumbaiah, C. Borchers, A.-C. Falhs, K. Holstein, N. Rummel, and V. Aleven. Teacher noticing and student learning in human-ai partnered classrooms: a multimodal analysis. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023, pp. 1042-1045*. International Society of the Learning Sciences, 2023.
  - [18] S. Karumbaiah, C. Borchers, T. Shou, A.-C. Falhs, P. Liu, T. Nagashima, N. Rummel, and V. Aleven. A spatiotemporal analysis of teacher practices in supporting student learning and engagement in an ai-enabled classroom. In *International Conference on Artificial Intelligence in Education*, pages 450–462. Springer, 2023.
  - [19] T. Kerry and C. A. Kerry. Differentiation: Teachers' views of the usefulness of recommended strategies in helping the more able pupils in primary and secondary classrooms. *Educational Studies*, 23(3):439–457, 1997.
  - [20] K. R. Koedinger, P. F. Carvalho, R. Liu, and E. A. McLaughlin. An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences*, 120(13):e222131120, 2023.
  - [21] Y. Lee and J. A. Nelder. Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(4):619–656, 1996.
  - [22] R. Liu and K. R. Koedinger. Towards reliable and valid measurement of individualized student parameters. *International Educational Data Mining Society*, 2017.
  - [23] H. Margolis\*. Increasing struggling learners' self-efficacy: What tutors can do and say. *Mentoring & Tutoring: Partnership in Learning*, 13(2):221–238, 2005.
  - [24] R. Martinez-Maldonado, J. Kay, K. Yacef, M.-T. Edbauer, and Y. Dimitriadis. Mtclassroom and mtdashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. 2013.
  - [25] Z. A. Pardos, R. S. Baker, M. O. San Pedro, S. M. Gowda, and S. M. Gowda. Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 117–124, 2013.
  - [26] J. W. Schofield, R. Eurich-Fulcer, and C. L. Britt. Teachers, computer tutors, and teaching: The artificially intelligent tutor as an agent for classroom change. *American Educational Research Journal*, 31(3):579–607, 1994.
  - [27] J. Stamper, S. Moore, C. Rose, P. Pavlik, K. Koedinger, et al. Learnsphere: A learning data and analytics cyberinfrastructure. *Journal of Educational Data Mining*, 16(1):141–163, 2024.
  - [28] M. Tissenbaum, C. Matuk, M. Berland, L. Lyons, F. Cocco, M. Linn, J. L. Plass, N. Hajny, A. Olsen, B. Schwendimann, et al. Real-time visualization of student activities to support classroom orchestration. Singapore: International Society of the Learning Sciences, 2016.