Teacher Noticing and Student Learning in Human-AI Partnered Classrooms: A Multimodal Analysis

Shamya Karumbaiah, University of Wisconsin-Madison, shamya.karumbaiah@wisc.edu
Conrad Borchers, Carnegie Mellon University, cborcher@cs.cmu.edu
Ann-Christin Falhs, Ruhr University Bochum, ann-christin.falhs@rub.de
Kenneth Holstein, Carnegie Mellon University, kjholste@cs.cmu.edu
Nikol Rummel, Ruhr University Bochum, nikol.rummel@rub.de
Vincent Aleven, Carnegie Mellon University, aleven@cs.cmu.edu

Abstract: Past research shows that teacher noticing matters for student learning, but little is known about the effects of AI-based tools designed to augment teachers' attention and sensemaking. In this paper, we investigate three multimodal measures of teacher noticing (i.e., gaze, deep dive into learning analytics in a teacher tool, and visits to individual students), gleaned from a mixed reality teacher awareness tool across ten classrooms. Our analysis suggests that of the three noticing measures, deep dive exhibited the largest association with learning gains when adjusting for students' prior knowledge and tutor interactions. This finding may indicate that teachers identified students most in need based on the deep dive analytics and offered them support. We discuss how these multimodal measures can make the constraints and effects of teacher noticing in human-AI partnered classrooms visible.

Introduction

Teacher noticing of noteworthy events in the classroom is argued to be a key teacher competency for effective pedagogical practice (Blömeke et al., 2015) and has been shown to matter for student learning and experience (Kersting et al., 2012). At the same time, research on human perception suggests that teachers may be limited in their attentional capacity when focusing on the multitudes of events happening in a classroom at a given point in time (Kahneman, 1973). Classrooms with artificially intelligent tutors (AI tutors) that collect moment-by-moment data on student learning and experience offer new opportunities for AI-supported teacher noticing. Through analytics derived from these data, teachers can react to classroom events that they otherwise might not have been aware of. Accordingly, AI tools designed to improve teachers' real-time awareness and sensemaking have been effective in improving students' learning (Holstein et al., 2018). However, we are yet to fully understand how teachers distribute their attention across students when provided with analytics that extend their current knowledge about classroom learning. These settings offer novel opportunities to study how teacher noticing and intervention relate to student learning and engagement with AI tutors. Besides being interesting in its own right, this kind of understanding could be helpful in designing better tools that support teachers' in-the-moment noticing and ultimately support reflection on their own practices. In this study, we relate three multimodal measures of teacher noticing to student learning in an AI classroom with a mixed reality teacher tool. We showcase how these measures can be used to better understand and quantify teacher noticing and its relationship to student learning.

Multimodal measures of teacher noticing in human-AI partnered classrooms

In this paper, we measure teacher noticing using multimodal data collected from a mixed reality teacher awareness tool. Lumilo is a smart glass system that sends real-time analytics about student learning and engagement (i.e., idle, rapid attempts, hint abuse, low or high local error rates, many errors after hints, hint avoidance, and unproductive persistence) to the teacher via indicator icons (Holstein et al., 2018; Holstein & Aleven, 2022). In addition to providing teachers with real-time analytics, the smart glasses are instrumented to gather multimodal measures of teacher interaction. We investigate three such measures: 1) the number of visual fixations teachers allocate to students (*gaze*), 2) how often teachers use the deep dive function of Lumilo to obtain more information on a student's progress in the AI tutor (*deep dive*), and 3) how often teachers visit students in person (*visit*).

According to the framework proposed by van Es and Sherin (2021), teacher noticing involves *attending* i.e., recognizing important aspects of classroom interactions and ignoring others, *interpreting* i.e., reasoning about what is observed using contextual knowledge and past experiences and *shaping* i.e., gathering additional information by constructing new interactions with the students while continuing to notice. We argue that our three multimodal measures (i.e., gaze, deep dive, visit) operationalize key components of van Es and

Sherin's (2021) noticing framework in a manner that applies to the targeted classroom scenario. First, visual attention in the form of gaze is unequally distributed among students; it tends to be focused particularly on students who exhibit undesirable behavior (Wolff et al., 2017; Yamamoto & Imai-Matsumura, 2013). Prior work also suggests that students with low academic performance are more often in the teacher's visual focus (Chaudhuri et al., 2022). Based on these findings, teachers' gaze represents a coarse-grained measure of teachers' visual focus to identify noteworthy events (akin to *attending* in van Es & Sherin, 2021). Second, to gather further insight into noteworthy events, the Lumilo mixed reality tool, which generates the data analyzed in the current study, allows the teacher to open a deep dive analytics screen. This screen shows in-depth analytics on the given student's progress and instructional needs (e.g., areas of struggle; Holstein & Aleven, 2022). These real-time analytics augment teacher sensing, particularly concerning *interpreting* noteworthy events (van Es & Sherin, 2021). Still, teachers must connect multimodal analytics with contextual knowledge (e.g., prior knowledge of the students) to draw meaningful conclusions and guide further action (Deunk et al., 2018; Holstein & Aleven, 2022). Third, we view the use of the deep dive function and visit as forms of *shaping*, as defined by van Es and Sherin (2021). Shaping helps teachers amend and confirm assumptions gathered via attending and interpreting by gaining access to additional information.

We posit that it is worth investigating teacher noticing through multimodal measures for three reasons. First, multimodality offers a quantified lens into noticing. Second, multimodality offers a more complete representation of teacher noticing facets through multiple measures compared to a single measure. Third, multimodality allows for a lens into different levels of noticing, relating to theoretical stages of teacher noticing (cf. van Es & Sherin, 2021). In our context, we conjecture that gaze fixations, deep dives analytics lookups, and visits can be understood as stages of increasingly focused noticing of student needs for attention and extra help. Whereas the teacher's gaze wanders through the room or can be directed specifically to a student, calling up the deep dive function might indicate (perhaps especially when it involves a student just gazed upon) a heightened form of teacher noticing as the teacher proactively seeks further information about a student. In the same vein, perhaps visit can be understood as an action resulting from the preceding noticing events (attending, interpreting) to gather more information from the students directly. More research is needed to understand how teachers use different forms of noticing and how they relate to student learning. Hence, the guiding question for our analysis is: *how do multimodal measures of teacher noticing relate to student learning with AI tutors?* We hypothesize that physical teacher visits represent the most salient mode of teacher noticing from a student's perspective and is, therefore, most strongly related to students' learning gains compared to gaze and deep dive.

Methods

We analyze previously collected data from an intervention study investigating the efficacy of the mixed reality teacher awareness tool Lumilo (see previous section; Holstein et al., 2018). Our study sample included 173 students from ten classes taught by six teachers. Between pre- and post-test assessments of students' skills in the relevant mathematics (i.e., equation solving), students practiced math with the linear equation tutor Lynnette while being monitored by their teacher and supported when necessary. Each student worked with Lynnette for approximately 60 minutes across two classroom sessions. Their problem-solving behavior was recorded in the form of time-stamped log data. Lynnette has been reported to significantly improve students' equation-solving abilities. Teacher noticing variables were exported via Microsoft Hololens (see previous section). They include the number of times the teacher looked at a particular student (i.e., gaze), how often they used the deep dive feature to gather in-depth insight into a particular student (i.e., deep dive) and how often teachers entered the physical proximity of a student, defined as within a radius of four feet (i.e., visit). If a teacher entered the proximity of multiple students simultaneously, proximity was coded for the student closest to the teacher. From the tutor log data, we aggregated the following student-level variables: (1) tutor interactions, such as the average time students take for correct, incorrect, and all steps when working with the AI tutor, (2) in-system performance, such as ratio of correct to incorrect attempts, and (3) engagement behaviors, such as tutor misuse, estimated using previously-developed machine learning models for the AI tutor (cf. Holstein et al., 2019). Grade level, prior student knowledge, and class size served as control variables. As our outcome, we analyzed the association of these variables with students' learning gain after working with the AI tutor. Learning gains were operationalized as the difference between normalized pre- and post-test scores.

Results

We investigate whether a teacher visit is most strongly related to learning gains among our three multimodal noticing measures as it could be the most salient mode of teacher noticing from a student perspective. We employ an automatic feature selection procedure (AIC-based backward feature selection) for a linear regression model of learning gain, adjusting for prior knowledge, grade level and student-tutor interactions. Contrary to our hypothesis, deep dive, not visit, had the strongest association with learning gains after controlling for contextual factors and students' tutor behavior (β = 0.19 [0.07, 0.31], p = .001; Table 2). Dive explained 3.1% of the variance in learning gains beyond the other variables featured in the selected model. As deep dive was the only noticing measure selected by our procedure, gaze and visit did not explain a significant amount of variance in learning gains beyond deep dive.

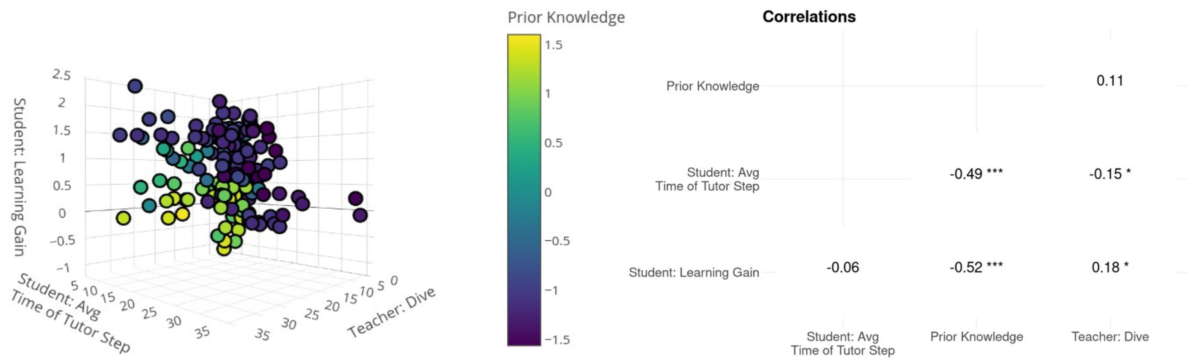
Table 2

Linear model of learning gain selected via backward search based on AIC with deep dive being the only noticing measure that explained variance in learning gains beyond control variables.

Predictors	***Estimates***	***CI***	***p***
Intercept	0.69	0.56 – 0.82	**<.001**
Deep Dive	0.19	0.07 – 0.31	**.001**
Avg Time Tutor Step	-0.51	-0.78 – -0.25	**<0.001**
Avg Time Correct Tutor Step	0.16	-0.05 – 0.36	.129
Avg Time Incorrect Tutor Step	0.17	0.05 – 0.28	**.005**
Avg Tutor Misuse Score	-0.07	-0.15 – 0.01	.107
# Idle Tutor Sequences	-0.13	-0.26 – 0.00	.055
Avg Length of Idle Tutor Sequences	-0.13	-0.23 – -0.02	**.018**
Avg Peaks of Idle Tutor Sequences	0.09	-0.03 – 0.21	.155
Avg Peaks of Struggle Sequences	-0.07	-0.17 – 0.03	.173
Prior Knowledge/Pre Test Score	-0.69	-0.82 – -0.55	**<.001**
Grade Level [7th]	-0.11	-0.54 – 0.33	.632
Grade Level [8th]	0.10	-0.11 – 0.31	.347
Class Size	0.02	-0.11 – 0.14	.792
Avg Tutor Session Length	0.00	-0.10 – 0.10	.999
Observations	173		
R^2 / R^2 adjusted	0.529 / 0.488		

Figure 1

*3D scatter plot (a) of the most predictive variables of our learning gain model, with students standardized prior knowledge as color scale (light dots indicating low and dark dots indicating high pre-test score). Correlation heatmap (b) of the same variables' intercorrelations with significant levels (***p<0.001 **p<0.01 *p<0.05).*



(a) (b)

Next, we check heterogeneous effects of teacher noticing across students by plotting the most associated noticing variable (i.e., deep dive), tutor variable (i.e., average time spent per tutor step), and control variable (i.e., prior knowledge) in Figure 1. We observe three main trends. First, we find a significant positive (albeit small) association between deep dive and learning gains and a significant negative association between the average time per tutor step and prior knowledge. Second, teachers performed significantly more deep dives on students who spent more time per tutor step, while students with longer tutor steps had significantly lower prior knowledge. Third, a small (non-significant) positive correlation was found between deep dive and prior knowledge.

Discussion

The current study investigates how multimodal teacher noticing measures relate to learning in a human-AI partnered classroom. We hypothesized that physical visits would be most strongly related to students' learning. Contrary to that hypothesis, deep dive had the strongest positive association with learning gains. This finding may indicate that teachers identified students in need based on the deep dive analytics and offered them support, potentially acting upon diagnosed struggles and difficulties students experienced. However, since visits were less strongly associated with learning than deep dives, the support prompted by deep dives would, apparently, not always be in the form of an actual visit, consistent with field observations that some teachers using Lumilo would frequently provide feedback to students from across the room, without physically visiting them (Holstein et al., 2018; 2019). Perhaps other purposes of teacher visits not immediately related to student support, for example, looking over a student's shoulder without interacting with the student, celebrating the successful completion of a problem, or other social interactions might dilute the current signal for teacher support for learning captured by our visit measure. Another potential explanation is that teacher visits were actually helpful, but that visits nonetheless did not correlate positively with learning gains because these visits were infrequent. As teachers tended to perform deep dives on students with higher prior knowledge, mechanisms other than support could also explain the lack of association between visit and learning. For example, teachers' selection of students to perform deep dives might have been influenced by top-down selection effects (i.e., decisions informed by teachers' prior knowledge of the student rather than noteworthy classroom behavior; Chaudhuri et al., 2022). From post-study interviews with teachers during data collection, we know that teachers report using the deep dive feature on multiple high-achieving students in a row to calibrate their assessment of class progress (Holstein et al., 2019). Thus, the use of deep dives may sometimes represent teachers' efforts to gather classroom-level data rather than sizing up the extent of a student's struggle. Future research may look at how deep dives with different purposes might be distinguished (e.g., their duration might be different), which may re-inform or help extend prior conceptualizations of teacher noticing. In conclusion, our multimodal measures offer novel lenses into the study of teacher noticing in classrooms with AI tutors that we believe would have implications to designing tools for teacher orchestration and reflection.

References

Blömeke, S., Gustafsson, J., & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13.

Chaudhuri, S., Muhonen, H., Pakarinen, E. & Lerkkanen, M.-K. (2022). Teachers' visual focus of attention in relation to students' basic academic skills and teachers' individual support for students: An eye-tracking study. *Learning and Individual Differences*, *98*, 102179.

Deunk, M. I., Smale-Jacobse, A. E., Boer, H. de, Doolaard, S., & Bosker, R. J. (2018). Effective differentiation Practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*, 31–54.

Holstein, K., & Aleven, V. (2022). Designing for human–AI complementarity in K-12 education. *AI Magazine*, *43*(2), 239–248.

Holstein, K., McLaren, B. M., & Aleven, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *Proceedings, AIED 2018* (Part 1, pp. 154–168). Cham: Springer.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics*, *6*(2).

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs: Prentice-Hall Inc.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589.

van Es, E. A., & Sherin, M. G. (2021). Expanding on prior conceptualizations of teacher noticing. *ZDM – Mathematics Education, 53*(1), 17–27.

Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. A. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education, 66*, 295-308.

Yamamoto, T., & Imai-Matsumura, K. (2013). Teachers' gaze and awareness of students' behavior: Using an eye tracker. *Comprehensive Psychology*, *2*, 01-IT.

Acknowledgments

We would like to thank NSF (Division of Information & Intelligent Systems award IIS-2119501) for sponsoring this project. The first three authors contributed equally to this manuscript. The authors would like to thank Kay Nam, Hanqing Liu, and Tzen-Hao Ng for their contributions to data preparation.