Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms

Kenneth Holstein(✉), Bruce M. McLaren, and Vincent Aleven

Carnegie Mellon University, Pittsburgh, PA 15213, USA
{kjholste,bmclaren,aleven}@cs.cmu.edu

Abstract. When used in K-12 classrooms, intelligent tutoring systems (ITSs) can be highly effective in helping students learn. However, they might be even more effective if designed to *work together* with human teachers, to amplify their abilities and leverage their complementary strengths. In the present work, we designed a wearable, real-time teacher awareness tool: mixed-reality smart glasses that tune teachers in to the rich analytics generated by ITSs, alerting them to situations the ITS may be ill-suited to handle. A 3-condition experiment with 286 middle school students, across 18 classrooms and 8 teachers, found that presenting teachers with real-time analytics about student learning, metacognition, and behavior had a positive impact on student learning, compared with both business-as-usual and classroom monitoring support without advanced analytics. Our findings suggest that real-time teacher analytics can help to narrow the gap in learning outcomes across students of varying prior ability. This is the first experimental study showing that real-time teacher analytics can enhance student learning. This research illustrates the promise of AIED systems that integrate human and machine intelligence to support student learning.

Keywords: Real-time analytics · Classroom orchestration
Teacher-in-the-loop · Human-AI hybrid · Intelligent tutors · Dashboards
Classroom evaluation

1 Introduction

When educational technologies are used in K-12 classrooms, human teachers play critical roles in mediating their effectiveness [33, 37, 41]. The term *classroom orchestration* has been widely used to describe the planning and real-time management of classroom activities [15]. Supporting teachers in orchestrating complex, but effective, technology-enhanced learning has been recognized as a critical research and design challenge for the learning sciences [16, 39, 43].

In recent years, several real-time teacher awareness tools have been designed and developed to address this challenge (e.g., [1, 5, 19, 30, 31, 40, 43]). These tools are often designed to augment teachers' "state awareness" during ongoing learning activities [39, 43], for example, by presenting teachers with real-time analytics on student knowledge, progress, metacognition, and behavior within educational software.

C. Penstein Rosé et al. (Eds.): AIED 2018, LNAI 10947, pp. 154–168, 2018.
https://doi.org/10.1007/978-3-319-93843-1_12

The design of such tools is frequently motivated by an assumption that enhanced teacher awareness will lead to improved teaching, and consequently, to improved student outcomes. Some prior work has found evidence for positive effects of real-time teacher analytics on student *performance* within educational software (e.g., [30]). Yet there is a paucity of empirical evidence that a teacher's use of real-time awareness tools (e.g., dashboards) can improve student *learning*, and scientific knowledge about the effects such tools have on teaching and learning is scarce [32, 39, 44].

In the present work, we investigate the effects of a real-time awareness tool for teachers working in K-12 classrooms using intelligent tutoring systems (ITSs), an important but underexplored area of AIED research. Intelligent tutors are a class of advanced learning technologies that provide students with step-by-step guidance during complex learning activities. ITSs have been found, in several meta-reviews, to significantly enhance student learning compared with other learning technologies or classroom instruction (e.g., [25]). When used in K-12 classrooms, ITSs allow students to work at their own pace, while also freeing up the teacher to spend more time working one-on-one with students [41]. A common intuition is that, in many situations, human teachers may be better suited to support students than ITSs alone (e.g., by providing socio-emotional support, supporting student motivation, or flexibly providing conceptual support when further problem-solving practice may be ineffective). Yet ITSs are not typically designed to work together with teachers, in real-time, to take advantage of these complementary strengths [7, 19, 35, 45]. ITSs might be even more effective if they were designed not only to support students directly, but also to amplify teachers' abilities to help their students (cf. [7, 21, 38]).

We present *Lumilo* [19]: mixed-reality smart glasses, co-designed with middle-school mathematics teachers, that tune teachers in to the rich analytics generated by ITSs (cf., [10]). By alerting teachers in real-time to situations the ITS may be ill-suited to handle on its own, *Lumilo* facilitates a form of mutual support or *co-orchestration* [35] between the human teacher and the AI tutor. We conduct an in-vivo experiment using *Lumilo* to investigate the effects of this form of teacher/AI co-orchestration on the ways teachers interact with students during in-school lab sessions with ITS software, and how, in turn, students learning processes and outcomes are affected. We test whether students measurably learn better when their teacher has access to real-time analytics from an ITS, compared to current practice with ITSs (where the teacher does not use a real-time awareness tool), and compared to a simpler form of *classroom monitoring* support [39], common in widely-used classroom-management systems (e.g., [11, 18, 26]).

2 Methods

2.1 Linear Equation Tutor

We investigate the effects of a teacher's use of a real-time awareness tool in the context of middle school classrooms using *Lynnette*, an ITS for linear equations. *Lynnette* is a rule-based Cognitive Tutor that was developed using the Cognitive Tutor Authoring Tools [3]. It has been used in several classroom studies, where it has been shown to significantly improve students' equation-solving ability [27, 28]. *Lynnette* provides step-by-step guidance, in the form of hints, correctness feedback, and error-specific messages as students tackle each problem in the software. It also adaptively selects problems for each student, using Bayesian Knowledge Tracing (BKT) to track individual students' knowledge growth, together with a mastery learning policy [13]. Students using *Lynnette* progress through five levels with equation-solving problems of increasing difficulty. These range from simple one-step equations at Level 1 (e.g., $x + 3 = 6$), to more complex, multi-step equations at Level 5 (e.g., $2(1 - x) + 4 = 12$).

2.2 Real-Time Teacher Awareness Tool

We created a real-time support tool for K-12 teachers who use ITSs in their classrooms. To this end, we adopted a participatory design approach [17] in which we directly involved teachers at each stage, from initial needs-finding [19, 21] to the selection and tuning of real-time analytic measures through iterative prototyping [19, 20]. The prototype that emerged from this iterative co-design process (described in greater detail in [19–21]), was a pair of mixed-reality smart glasses called *Lumilo*.

Lumilo tunes teachers in to the rich analytics that ITSs generate: It presents real-time indicators of students' current learning, metacognitive, and behavioral "states", projected in the teacher's view of the classroom (Fig. 1, left). The use of transparent smart glasses allows teachers to keep their heads up and focused on the classroom, enabling them to continue monitoring important signals that may not be captured by the tool alone (e.g., student body language and looks of frustration [19, 21]). The smart glasses provide teachers with a *private* view of actionable, real-time information about their students, embedded throughout the classroom environment, thus providing many of the advantages of ambient and distributed classroom awareness tools (e.g., [1, 5]), without revealing sensitive student data to the entire class [5, 19].

Over the course of the design process, *Lumilo's* information displays shifted towards strongly minimalistic designs (with progressive disclosure of additional analytics only upon a teacher's request), in accordance with the level of information teachers desired and could handle in fast-paced classroom environments. *Lumilo* presents mixed-reality displays of three main types, visible through the teacher's glasses: student-level indicators, student-level "deep-dive" screens, and class-level summaries (as shown in Fig. 1). Student-level indicators and class-level summaries are always visible to the teacher by default, at a glance. Student-level indicators display above corresponding students' heads (based on teacher-configurable seating charts), and class-level summaries can display at teacher-configurable locations throughout the classroom [19]. As shown in Fig. 1(bottom-left), if a teacher glances at a student's indicator, *Lumilo* automatically displays a brief elaboration about the currently displayed indicator symbol (i.e., how long the alert has been active and/or a brief explanation of *why* the alert is showing). If no indicators are currently active for a student, *Lumilo* displays a faint circular outline above that student (Fig. 1, top-left). If a teacher clicks on a student's indicator (using either a handheld clicker or by making a 'tap' gesture in mid-air), *Lumilo* displays "deep-dive" screens for that student. As shown in Fig. 1(right), these screens include a "Current Problem" display, which supports remote *monitoring*, showing a live feed of a student's work on their current problem. Each problem step in this feed is annotated with the number of hint requests and errors the student has made on that step. In classroom observations, we have found that because *Lumilo* enables monitoring of student activities from a distance (i.e., across the room), teachers using *Lumilo* often interleave help across students: While helping one student at that student's seat, the teacher might provide quick guidance to a struggling student across the room (Fig. 2, right).

Fig. 1. Teacher's point-of-view while using *Lumilo*. Top row: illustrative mock-ups; Bottom row: screenshots captured through *Lumilo* (taken after the end of a class session, to protect student privacy) [19]. Left: Teacher's default view of the class through *Lumilo*. Right: Deep-dive screens that pop-up if a teacher 'clicks' on a student's indicator.

The deep-dive screens also include an "Areas of Struggle" screen, which displays the three skills for which a student has the lowest probability of mastery. For each skill shown in "Areas of Struggle", the student's estimated probability of mastery is displayed, together with a concrete example of an error the student has made on a recent practice opportunity for the skill. In addition, in the current study, a class-level summary display was available to the teacher: the "Low Mastery, High Practice" display (illustrated on the left, in the top row images of Fig. 1). This display shows the three skills that the fewest students in the class have mastered (according to BKT), at a given point in the class session, out of those skills that many students in the class have already had opportunities to practice within the software [19].

The student indicators displayed by *Lumilo* (Fig. 2, left) are ideas that were generated by teachers in our design studies [19, 21] and implemented using established

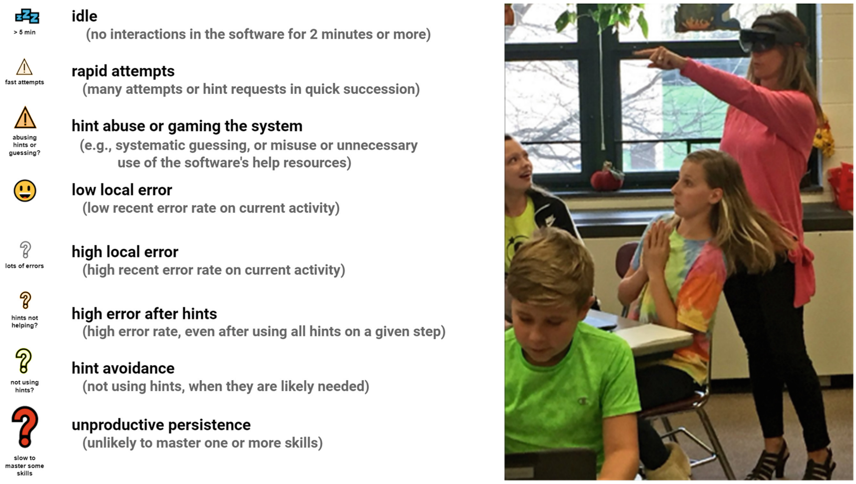


Fig. 2. Left: Indicators displayed by *Lumilo* [20]. Right: Teacher using *Lumilo* during class.

student modeling methods (e.g., [2, 4, 9, 13, 14, 23]). The analytic measures and their corresponding visual alerts were iteratively refined based on prototyping feedback from teachers [19], as well as causal data mining of teacher and student process data from classroom pilots using *Lumilo* [20]. The resulting prototype updates real-time student indicators based on the outputs of sensor-free detectors, including detectors of student hint abuse and hint avoidance [2, 4], gaming-the-system [8], rapid/non-deliberate step attempts or hint requests [2], and unproductive persistence or "wheel-spinning" [9, 23]. In addition, *Lumilo* indicates when a student has been idle for two minutes or more and may be off-task (cf. [6]), when a student has been exhibiting a particularly "low" or "high" recent error rate (less than 30% or greater than 80% correct within the student's most recent 10 attempts) (cf. [23, 34]), or when a student is making errors on a given problem-solving step, despite having already exhausted all tutor-provided hints for that step [2]. By directing teachers' attention, in real-time, to situations the ITS may be ill-suited to handle, *Lumilo* is designed to facilitate productive mutual support or *co-orchestration* [35] between the teacher and the ITS, by leveraging the complementary strengths of each (cf. [22, 35, 38, 45]).

2.3 Experimental Design, Participants, and Procedure

In this study, we investigated the hypothesis that real-time teacher/AI co-orchestration, supported by real-time analytics from an ITS, would enhance student learning compared with both (a) business-as-usual for an ITS classroom, and (b) classroom *monitoring support* without advanced analytics (a stronger control than (a), as described below).

To test these hypotheses, we conducted a 3-condition experiment with 343 middle school students, across 18 classrooms, 8 teachers, and 4 public schools (each from a different school district) in a large U.S. city and surrounding areas. All participating teachers had at least 5 years of experience teaching middle school mathematics and had previously used an ITS in their classroom. The study was conducted during the first half of the students' school year, and none of the classes participating in this study had previously covered equation-solving topics beyond simple one-step linear equations (e.g., $x - 2 = 1$).

Classrooms were randomly assigned to one of three conditions, stratified by teacher. In the Glasses+Analytics condition, teachers used the full version of *Lumilo*, including all displays described above. In the business-as-usual (noGlasses) condition, teachers did not wear *Lumilo* during class, and thus did not have access to real-time analytics. We also included a third condition (Glasses) in which teachers used a reduced version of *Lumilo* with only its monitoring functionality (i.e., without any of its advanced analytics). This condition was included because prior empirical findings suggest that students' mere awareness that a teacher is monitoring their activities within an ITS may have a significant effect on student learning (e.g., by discouraging, and thus decreasing the frequency of maladaptive learning behaviors such as gaming-the-system) [22, 42]. In the Glasses condition, teachers only retained the ability to "peek" at students' screens from any location in the classroom, using the glasses (although without the line-by-line annotations present in *Lumilo*'s "Current Problem" screen). All of *Lumilo*'s student indicators were replaced by a single, static symbol (a faint circular outline) that did not convey any information about the student's state. Further, the "Areas of Struggle" deep dive screens and the class-level displays were hidden. Our aim in providing this stripped-down version *of Lumilo* was to encourage teachers to interact with the glasses, thereby minimizing differences in students' perceptions between the Glasses+Analytics and Glasses conditions. The Glasses condition bears some similarity to standard classroom monitoring tools, which enable teachers to peek at student screens on their own desktop or tablet display (e.g., [11, 18, 26]).

All teachers participated in a brief training session before the start of the study. Teachers were first familiarized with *Lynnette*, the tutoring software that students would use during the study. In the Glasses+Analytics and Glasses conditions, each teacher also participated in a brief (30-min) training with *Lumilo* before the start of the study. In this training, teachers practiced interacting with two versions of the glasses (Glasses and Glasses+Analytics) in a simulated classroom context. At the end of this training, teachers were informed that, for each of their classes, they would be assigned to use one or the other of these two designs.

Classrooms in each of the three conditions followed the same procedure. In each class, students first received a brief introduction to *Lynnette* from their teacher. Students then worked on a computer-based pre-test for approximately 20 min, during which time the teacher provided no assistance. Following the pretest, students worked with the tutor for a total of 60 min, spread across two class sessions. In all conditions, teachers were encouraged to help their students as needed, while they worked with the tutor. Finally, students took a 20-min computer-based post-test, again without any assistance from the teacher. The pre- and posttests focused on procedural knowledge of equation solving. We used two isomorphic test forms that varied only by the specific numbers used in equations. The tests forms were assigned in counterbalanced order across pre- and post-test. The tests were graded automatically, with partial credit assigned for intermediate steps in a student's solution, according to *Lynnette*'s cognitive model.

In the Glasses and Glasses+Analytics conditions, we used *Lumilo* to automatically track a teacher's physical position within the classroom (cf. [36]), relative to each student, moment-by-moment (leveraging *Lumilo*'s indicators as mixed-reality proximity sensors [19, 20]). Teacher time allocation was recorded per student as the cumulative time (in seconds) a teacher spent within a 4-ft radius of that student (with ties resolved by relative proximity). Given our observation that teachers in both of these conditions frequently provided assistance remotely (i.e., conversing with a student from across the room, while monitoring her/his activity using the glasses), teacher time was also accumulated for the duration a teacher spent peeking at a student's screen via the glasses. In the noGlasses condition, since teachers did not wear *Lumilo*, time allocation was recorded via live classroom coding (using the *LookWhosTalking* tool [29]) of the target (student) and duration (in seconds) of each teacher visit. In addition to test scores and data on teacher time allocation, we analyzed tutor log data to investigate potential effects of condition on students' within-software behaviors.

3 Results

Fifty-seven students were absent for one or more days of the study and were excluded from further analyses. We analyzed the data for the remaining 286 students. Given that the sample was nested in 18 classes, 8 teachers, and 4 schools, and that the experimental intervention was applied at the class level, we used hierarchical linear modeling (HLM) to analyze student learning outcomes. 3-level models had the best fit, with students (level 1) nested in classes (level 2), and classes nested in teachers (level 3). We used class track (low, average, or high) as a level-2 covariate. Both 2-level models, (with students nested in classes) and 4-level models (with teachers nested in schools) had worse fits according to both AIC and BIC, and 4-level models indicated little variance on the school level. We report r for effect size. An effect size r above 0.10 is conventionally considered small, 0.3 medium, and 0.5 large [12].

Effects on Student Learning. To compare student learning outcomes across experimental conditions, we used HLMs with test score as the dependent variable, and test type (pretest/posttest, with pretest as the baseline value) and experimental condition as independent variables (fixed effects). For each fixed effect, we included a term for each comparison between the baseline and other levels of the variable. For comparisons between the Glasses+Analytics and noGlasses conditions, we used noGlasses as the condition baseline. Otherwise, we used Glasses as the baseline.

Across conditions, there was a significant gain between student pretest and posttest scores ($t(283) = 7.673, p = 2.74 * 10^{-13}, r = 0.26$, 95% CI [0.19, 0.34]), consistent with results from prior classroom studies using Lynnette [27, 28], which showed learning gain effect size estimates ranging from $r = 0.25$ to $r = 0.64$. Figure 3 shows pre-post learning gains for each condition. There was a significant positive interaction between student pre/posttest and the noGlasses/Glasses+Analytics conditions ($t(283) = 5.897$,

$p = 1.05 * 10^{-8}$, $r = 0.21$, 95% CI [0.13, 0.28]), supporting the hypothesis that real-time teacher/AI co-orchestration, supported by analytics from an ITS, would enhance student learning compared with business-as-usual for ITS classrooms.

Decomposing this effect, there was a significant positive interaction between student pre/posttest and the noGlasses/Glasses conditions ($t(283) = 3.386$, $p = 8.08 * 10^{-4}$, $r = 0.13$, 95% CI [0.02, 0.23]), with a higher learning gain slope in the Glasses condition, indicating that relatively minimal classroom monitoring support, even without advanced analytics, can positively impact learning. In addition, there was a significant positive interaction between student pre/posttest and the Glasses/Glasses+Analytics conditions ($t(283) = 2.229$, $p = 0.027$, $r = 0.11$, 95% CI [0.02, 0.20]), with a higher slope in the Glasses+Analytics condition than in the Glasses condition, supporting our hypothesis that real-time teacher analytics would enhance student learning, above and beyond any effects of monitoring support alone (i.e., without advanced analytics).

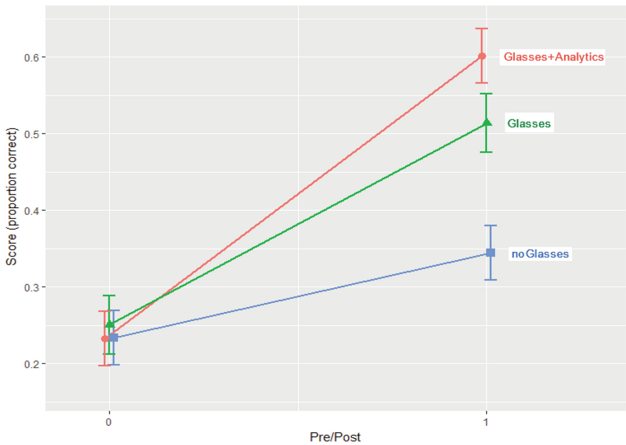
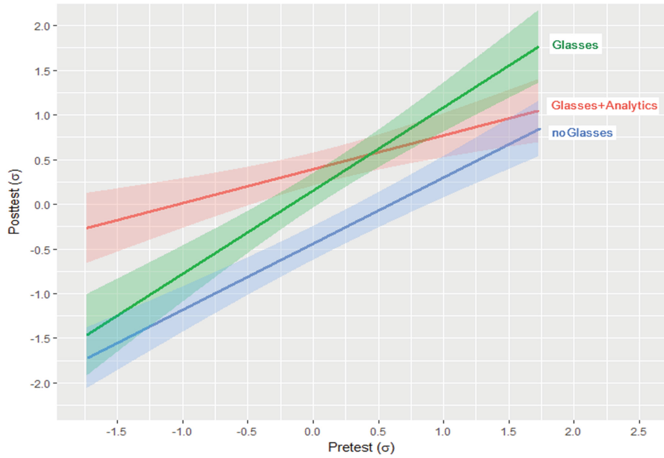


Fig. 3. Student pre/post learning, by experimental condition. Error bars indicate standard error.

Aptitude-Treatment Interactions on Student Learning. We next investigated how the effects of each condition might vary based on students' prior domain knowledge. *Lumilo* was designed to help teachers quickly identify students who are currently struggling (unproductively) with the ITS, so that they could provide these students with additional, on-the-spot support. If *Lumilo* was successful in this regard, we would expect to see an aptitude-treatment interaction, such that students coming in with lower prior domain knowledge (who are more likely to struggle) would learn more when teachers had access to *Lumilo's* real-time analytics [19, 20].

We constructed an HLM with posttest as the dependent variable and pretest and experimental condition as level-1 covariates, modeling interactions between pretest and condition. Figure 4(top) shows student posttest scores plotted by pretest scores (in standard deviation units) for each of the three conditions. As shown, students in the Glasses condition learned more overall, compared with the noGlasses condition, but the disparity in learning outcomes across students with varying prior domain knowledge remained the same. For students in the Glasses+Analytics condition, the posttest by pretest curve was flatter, with lower pretest students learning considerably more than in the other two conditions. There was no significant interaction between noGlasses/Glasses and student pretest. However, there were significant negative interactions between student pretest scores and noGlasses/Glasses+Analytics ($t(46) = -2.456$, $p = 0.018$, $r = -0.15$, 95% CI [−0.26, −0.03]) and Glasses/Glasses+Analytics ($t(164) = -2.279$, $p = 0.024$, $r = -0.16$, 95% CI [−0.27, −0.05]), suggesting that a teacher's use of real-time analytics may serve as an equalizing force in the classroom.



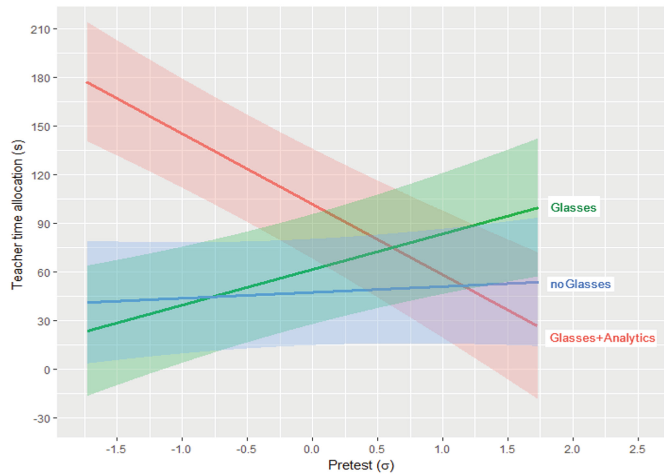


Fig. 4. Student posttest scores (top) and teacher attention allocation (bottom), plotted by student pretest scores, for each experimental condition. Shaded regions indicate standard error.

Effects on Teacher Time Allocation. As an additional way of testing whether the real-time analytics provided by *Lumilo* had their intended effect, we fit an HLM with teacher time allocation, per student, as the dependent variable, and student pretest score, experimental condition, and their interactions as fixed effects. Figure 4 (bottom) shows teacher time, plotted by student pretest, for each condition. As shown, in the Glasses+Analytics condition, teachers tended to allocate considerably more of their time to students with lower prior domain knowledge, compared to the other conditions. There was no significant main effect of noGlasses/Glasses on teacher time allocation ($t(211) = 0.482$, $p = 0.63$, $r = 0.03$, 95% CI [0, 0.14]), nor a significant interaction with pretest. However, there were significant main effects of noGlasses/Glasses+Analytics ($t(279) = 2.88$, $p = 4.26 * 10^{-3}$, $r = 0.17$, 95% CI [0.06, 0.28]) and Glasses/Glasses+Analytics ($t(278) = 2.02$, $p = 0.044$, $r = 0.12$, 95% CI [0.01, 0.23]) on teacher time allocation. In addition, there were significant negative interactions between student pretest and noGlasses/Glasses+Analytics ($t(279) = -2.88$, $p = 4.28 * 10^{-3}$, $r = -0.17$, 95% CI [−0.28, −0.05]) and Glasses/Glasses+Analytics ($t(275) = -3.546$, $p = 4.62 * 10^{-4}$, $r = -0.23$, 95% CI [−0.33, −0.11]).

We also investigated how teachers' relative time allocation across students may have been driven by the real-time analytics presented in the Glasses+Analytics condition. Specifically, we examined whether and how teacher time allocation varied across conditions, based on the frequency with which a student exhibited each of the within-tutor behaviors/states detected by *Lumilo* (i.e., *Lumilo*'s student indicators, described in Sect. 2.3). We constructed HLMs with teacher time allocation as the dependent variable, and the frequency of student within-tutor behaviors/states, experimental condition, and their interactions as fixed effects. Row 3 of Table 1 shows relationships between student within-tutor behaviors/states and teacher time allocation across students, for the Glasses+Analytics vs. noGlasses (GA v. nG) comparison. As shown, teachers' time allocation across students appears to have been influenced by *Lumilo*'s real-time indicators. Compared with business-as-usual (Row 3, Table 1), teachers in the Glasses+Analytics condition spent significantly *less* time attending to students who frequently exhibited low local error, and significantly *more* time attending to students who frequently exhibited undesirable behaviors/states detected by *Lumilo*, such as unproductive persistence (or "wheel-spinning").

Table 1. Estimated effects of condition (rows) on teachers' allocation of time to students exhibiting each within-tutor behavior/state (columns). Cells report estimated effect sizes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $0.05 \leq p < 0.07$

	High local error	Hint abuse or gaming	Hint avoidance	High error after hints	Idle	Low local error	Rapid attempts	Unproductive persistence ("wheel-spinning")
G v. nG	n.s.	n.s.	n.s.	n.s.	n.s.	0.13~	n.s.	n.s.
GA v. G	**0.20***	**0.17***	**0.19***	**0.18***	**0.22****	**−0.51*****	n.s.	**0.35*****
GA v. nG	**0.16****	0.10~	**0.14***	0.11~	**0.17****	**−0.23*****	n.s.	**0.24*****

Rows 1 and 2 of Table 1 show estimates for Glasses vs. noGlasses (G v. nG) and Glasses+Analytics vs. Glasses (GA v. G), respectively. As shown, there were no significant differences in teacher time allocation due to the introduction of the glasses themselves, suggesting *Lumilo*'s overall effects on teacher time allocation may result primarily from teachers' use of the advanced analytics presented in the GA condition.

Effects of Classroom Monitoring Support and Real-Time Teacher Analytics on Student-Level Processes. To investigate potential effects of experimental condition on the frequency of student within-tutor behaviors and learning states detected by *Lumilo*, we constructed HLMs with students' within-tutor behaviors/states as the dependent variable, and pretest score and experimental condition as fixed effects. Row 3 of Table 2 shows estimated effects of classroom condition on the frequency of student within-tutor behaviors/states, for Glasses+Analytics vs. noGlasses (GA v. nG).

Table 2. Estimated effects of condition (rows) on the frequency of student within-tutor behaviors/states (columns): *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $\sim$ $0.05 \leq p < 0.07$

	High local error	Hint abuse or gaming	Hint avoidance	High error after hints	Idle	Low local error	Rapid attempts	Unproductive persistence ("wheel-spinning")
G v. nG	**−0.36 ****	**−0.21****	**−0.32****	n.s.	**0.23***	**0.34*****	n.s.	n.s.
GA v. G	−0.12~	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	−0.20~
GA v. nG	**−0.47*****	**−0.28****	**−0.41*****	**−0.30*****	**0.26***	**0.42***	**−0.34****	**−0.15***

Compared with business-as-usual, students in the Glasses+Analytics condition exhibited less hint avoidance or gaming/hint abuse, were less frequently detected as unproductively persisting or making rapid consecutive attempts in the tutoring software and exhibited less frequent high local error. In addition, students in the Glasses+Analytics condition were more frequently idle in the software, and more frequently exhibited low local error. Row 1 of Table 2 suggests that that the introduction of the glasses, even without real-time teacher analytics, may have had a considerable influence on students' behavior within the software. By contrast, there were no significant differences between the Glasses+Analytics and Glasses conditions. These results suggest that, despite the ostensible positive effects of real-time teacher analytics on student learning outcomes, some of the *largest* effects of *Lumilo* on students' within-tutor behavior may result primarily from teachers' use of the monitoring support provided in the Glasses condition, rather than from a teachers' use of advanced analytics.

4 Discussion, Conclusions, and Future Work

We conducted a 3-condition classroom experiment to investigate the effects of a real-time teacher awareness tool on student learning in ITS classrooms. Our findings indicate that teachers' use of *Lumilo*, a real-time awareness tool, resulted in higher learning gains with the ITS. In addition, presenting teachers with real-time analytics about student learning, metacognition, and behavior at a glance had a positive impact on student learning with the ITS, above and beyond the effects of monitoring support alone (without any advanced analytics). The real-time analytics provided by *Lumilo* appear to have served as an equalizing force in the classroom: driving teachers' time towards students of lower prior ability and narrowing the gap in learning outcomes between students with higher and lower prior domain knowledge.

Interestingly, part of *Lumilo's* overall effect on student learning appears to be attributable to monitoring support alone. Follow-up correlational analyses suggested that a teacher's use of the glasses, with monitoring support (i.e., support for peeking at a student's screen remotely), but without advanced analytics, may reduce students' frequency of maladaptive learning behaviors (such as gaming/hint-abuse) without significantly influencing teachers' time allocation across students. These results suggest that the observed learning benefits of monitoring support may be due to a motivational effect, resulting from students' awareness that a teacher is monitoring their activities in the software (cf. [22, 42]), and/or due to a novelty effect. It may also be that the monitoring support provided in the Glasses condition had a positive effect on teacher behavior that is not reflected in the way they distributed their time across students (e.g., an effect upon teachers' verbal or non-verbal communication). Future work is needed to tease apart these explanations.

Although much prior work has focused on the design, development, and evaluation of teacher analytics tools, very few studies have evaluated effects on student learning [24, 32, 39, 44]). The current study is the first experimental study to demonstrate that real-time teacher analytics can enhance students' learning outcomes, within or outside the area of AIED and intelligent tutoring systems.

We see several exciting directions for future work. The current study involved teachers with at least five years of mathematics teaching experience. However, our prior design work with teachers indicated that less-experienced teachers may often struggle to generate effective on-the-spot help, in response to real-time analytics from an ITS [19, 22]. Thus, a promising direction for future design research is to investigate differences in needs for real-time support across teachers with varying levels of experience. In addition, while the current study was conducted over a single week of class time, future longitudinal studies may shed light on whether and how the effects of real-time teacher analytics and monitoring support may evolve over longer-term use (c.f. [32]). More broadly, an exciting direction for future work is to better understand and characterize the complementary strengths of human and automated instruction, to explore how they can most effectively be combined (cf. [21, 35, 38]).

In sum, this research illustrates the potential of AIED systems that integrate human and machine intelligence to support student learning. In addition, this work illustrates that the kinds of analytics already generated by ITSs, using student modeling techniques originally developed to support adaptive tutoring behavior, appear to provide a promising foundation for real-time teacher awareness tools.

Acknowledgements. This work was supported by NSF Award #1530726, and IES Grant R305B150008 to CMU. The opinions expressed are those of the authors and do not represent the views of NSF, IES or the U.S. ED. Special thanks to Gena Hong, Octav Popescu, Jonathan Sewall, Mera Tegene, Cindy Tipper, and all participating students and teachers.

References

1. Alavi, H.S., Dillenbourg, P.: An ambient awareness tool for supporting supervised collaborative problem solving. IEEE TLT **5**, 264–274 (2012)
2. Aleven, V.: Help seeking and intelligent tutoring systems: theoretical perspectives and a step towards theoretical integration. In: Azevedo, R., Aleven, V. (eds.) International Handbook of Metacognition and Learning Technologies, pp. 311–335. Springer, New York (2013). https://doi.org/10.1007/978-1-4419-5546-3_21
3. Aleven, V., McLaren, B.M., Sewall, J., Koedinger, K.R.: A new paradigm for intelligent tutoring systems: example-tracing tutors. IJAIED **19**(2), 105–154 (2009)
4. Aleven, V., Roll, I., McLaren, B.M., Koedinger, K.R.: Help helps, but only so much: research on help seeking with intelligent tutoring systems. IJAIED **26**, 205–223 (2016)
5. Alphen, E.V., Bakker, S.: Lernanto: using an ambient display during differentiated instruction. In: CHI EA (2016)
6. Baker, R.S.: Modeling and understanding students off-task behavior in intelligent tutoring systems. In: CHI, pp. 1059–1068 (2007)
7. Baker, R.S.: Stupid tutoring systems, intelligent humans. IJAIED **26**(2), 600–614 (2016)
8. Baker, R.S., Corbett, A.T., Roll, I., Koedinger, K.R.: Developing a generalizable detector of when students game the system. UMUAI **18**(3), 287–314 (2008)
9. Beck, J.E., Gong, Y.: Wheel-spinning: students who fail to master a skill. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 431–440. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_44
10. Bull, S., Kay, J.: Open learner models. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems, pp. 301–322. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14363-2_15
11. Chromebook Management Software for Schools. https://www.goguardian.com/
12. Cohen, J.: A power primer. Psychol. Bull. **112**(1), 155–159 (1992)
13. Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. UMUAI **4**(4), 253–278 (1995)
14. Desmarais, M.C., Baker, R.S.: A review of recent advances in learner and skill modeling in intelligent learning environments. UMUAI **22**(1–2), 9–38 (2012)
15. Dillenbourg, P., Jermann, P.: Technology for classroom orchestration. In: Khine, M., Saleh, I. (eds.) New Science of Learning, pp. 525–552. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-5716-0_26
16. Dillenbourg, P.: Trends in classroom orchestration. STELLAR **1**, 5 (2011)
17. Hanington, B., Martin, B.: Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport (2012)
18. Hapara | Making Learning Visible. https://hapara.com/
19. Holstein, K., Hong, G., Tegene, M., McLaren, B. M., Aleven, V.: The classroom as a dashboard: co-designing wearable cognitive augmentation for K-12 teachers. In: LAK, pp. 79–88. ACM (2018)
20. Holstein, K., McLaren, B.M., Aleven, V.: Informing the design of teacher awareness tools through causal alignment analysis. In: ICLS (in press)

21. Holstein, K., McLaren, B.M., Aleven, V.: Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In: LAK, pp. 257–266. ACM (2017)
22. Holstein, K., McLaren, B.M., Aleven, V.: SPACLE: investigating learning across virtual and physical spaces using spatial replays. In: LAK, pp. 358–367. ACM (2017)
23. Kai, S., Almeda, V.A., Baker, R.S., Shechtman, N., Heffernan, C., Heffernan, N.: Modeling wheel-spinning and productive persistence in skill builders. In: JEDM (in press)
24. Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., Soffer Goldstein, D.: Estimating the effect of web-based homework. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 824–827. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_122
25. Kulik, J.A., Fletcher, J.D.: Effectiveness of intelligent tutoring systems: a meta-analytic review. RER **86**(1), 42–78 (2016)
26. LanSchool Classroom Management Software. https://www.lenovosoftware.com/lanschool
27. Long, Y., Aleven, V.: Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 219–228. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_23
28. Long, Y., Aleven, V.: Gamification of joint student/system control over problem selection in a linear equation tutor. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 378–387. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07221-0_47
29. LookWhosTalking. bitbucket.org/dadamson/lookwhostalking
30. Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J.: MTFeedback: providing notifications to enhance teacher awareness of small group work in the classroom. IEEE TLT **8**(2), 187–200 (2015)
31. Mavrikis, M., Gutierrez-Santos, S., Poulovassilis, A.: Design and evaluation of teacher assistance tools for exploratory learning environments. In: LAK, pp. 168–172. ACM (2016)
32. Molenaar, I., Knoop-van Campen, C.: Teacher dashboards in practice: usage and impact. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) EC-TEL 2017. LNCS, vol. 10474, pp. 125–138. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66610-5_10
33. Nye, B.D.: Barriers to ITS adoption: a systematic mapping study. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 583–590. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07221-0_74
34. Pelánek, R., Řihák, J.: Experimental analysis of mastery learning criteria. In: UMAP, pp. 156–163. ACM (2017)
35. Prieto, L.P.: Supporting orchestration of blended CSCL scenarios in distributed learning environments. Unpublished doctoral thesis (2012)
36. Prieto, L.P., Sharma, K., Dillenbourg, P., Jesús, M.: Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In: LAK, pp. 148–157. ACM (2016)
37. Ritter, S., Yudelson, M., Fancsali, S.E., Berman, S.R.: How mastery learning works at scale. In: L@S, pp. 71–79. ACM (2016)
38. Ritter, S., Yudelson, M., Fancsali, S., Berman, S.R.: Towards integrating human and automated tutoring systems. In: EDM, pp. 626–627 (2016)
39. Rodríguez-Triana, M.J., Prieto, L.P., Vozniuk, A., Boroujeni, M.S., Schwendimann, B.A., Holzer, A., Gillet, D.: Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. IJTEL **9**(23), 126–150 (2017)

40. Segal, A., Hindi, S., Prusak, N., Swidan, O., Livni, A., Palatnic, A., Schwarz, B., Gal, Y.: Keeping the teacher in the loop: technologies for monitoring group learning in real-time. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 64–76. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61425-0_6
41. Schofield, J.W.: Computers and Classroom Culture. University Press, Cambridge (1995)
42. Stang, J.B., Roll, I.: Interactions between teaching assistants and students boost engagement in physics labs. Phys. Rev. Phys. Educ. Res. **10**(2), 020117 (2014)
43. Tissenbaum, M., Matuk, C.: Real-time visualization of student activities to support classroom orchestration. In: ICLS, pp. 1120–1127 (2016)
44. Xhakaj, F., Aleven, V., McLaren, B.M.: Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson planning, lessons and student learning. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) EC-TEL 2017. LNCS, vol. 10474, pp. 315–329. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66610-5_23
45. Yacef, K.: Intelligent teaching assistant systems. In: ICCE, pp. 136–140. IEEE (2002)