Towards Teacher-AI Hybrid Systems

Kenneth Holstein
Human-Computer Interaction Institute, Carnegie Mellon University
kjholste@cs.cmu.edu

ABSTRACT: AI-powered educational software, such as intelligent tutoring systems (ITSs), is commonly designed to enhance student learning. However, such software is not typically designed to effectively collaborate with human teachers. The present work explores how AI and teachers might leverage and amplify one another's complementary strengths to achieve outcomes greater than either could achieve alone. Key contributions of this research, so far, include (1) an initial, broad exploration of K-12 teachers' needs and desires for real-time support in ITS classrooms; (2) the first exploration of the affordances of smart glasses to support orchestration of personalized classrooms, resulting in a prototype called Lumilo; (3) a new prototyping method for real-time teacher support tools; and (4) a classroom experiment evaluating the effects of teacher/AI co-orchestration, supported by Lumilo, on teacher and student behavior and student learning outcomes in classrooms using ITSs.

Keywords: teaching analytics, classroom orchestration, cognitive augmentation, human-machine intelligence, human-in-the-loop, teachers, explainability, design methods, co-design

1 INTRODUCTION

To facilitate more personalized learning, AI-powered educational software is increasingly being used in K-12 classrooms (Pane, Steiner, Baird, Hamilton, & Pane, 2017). Yet teachers struggle with what their new roles in such classrooms should be (Holstein, McLaren, & Aleven, 2017b). One class of AI-powered educational software, intelligent tutoring systems (ITSs), have been shown through several meta-analyses to significantly enhance student learning when used in classrooms, compared with other forms of educational technology and traditional classroom instruction (e.g., Kulik & Fletcher, 2016). These systems provide step-by-step feedback and guidance to students – tailoring instruction to individual learners as they work through problem-solving activities at their own pace. In turn, ITSs also free up the teacher to provide more one-on-one support to students who may benefit from it the most. Yet orchestrating personalized learning also poses unique challenges for teachers, who are tasked with monitoring classes working on a range of educational activities, and prioritizing help across students given limited time (Holstein, McLaren, & Aleven, 2017b).

Over a decade ago, Yacef proposed a reframing of intelligent tutoring systems as "intelligent teaching assistants" (ITAs): systems designed with the joint objectives of helping human teachers teach and helping students learn (rather than only the latter of these objectives, as is typical of ITSs) (Yacef, 2002). Other researchers have since proposed similar directions, with a focus on optimizing student learning by leveraging complementary strengths of human and AI instruction (e.g., Ritter, Fancsali, Berman, & Yudelson, 2014). That is, ITSs might be more effective if they could adaptively enlist the help of human teachers (c.f. Kamar, 2016), in situations teachers may be better suited to handle. While there has been some work on real-time teacher support tools for ITS classrooms since the vision of ITAs was introduced, little work has explored teachers' actual *needs* and *desires* for such support, or how human instruction might most effectively be combined with AI instruction.

In my current and proposed work, I investigate how teachers might best be supported by AI, how AI might best be supported by teachers, and how AI and human teachers might leverage one another's complementary strengths to achieve outcomes greater than either could achieve alone.

2 PRIOR WORK

2.1 Exploring K-12 teachers needs and desires for real-time analytics

Whereas real-time support tools for teachers, such as dashboards, have become popular with many learning technologies (e.g., Martinez-Maldonado, Clayphan, Yacef, & Kay, 2015; Mavrikis, Gutierrez-Santos, & Poulovassilis, 2016), we are not aware of projects (in academic research or in industry) that have conducted a broad investigation of teachers' actual *needs* for real-time support (i.e., one that is not tied to an existing prototype or current technical feasibility, such as the current availability of data or measurement techniques (Holstein, McLaren, et al., 2017b; Rodriguez-Triana et al., 2017)). Furthermore, work on real-time support tools for personalized classrooms, more broadly, has tended to focus on designing tools for *university-level* instructors.

To better understand K-12 teachers' information needs for real-time analytics, my colleagues and I conducted a series of interviews and design studies with ten middle school math teachers (across 5 schools and 5 school districts in Pittsburgh and surrounding areas). For example, in a generative card sorting exercise, we asked teachers what "superpowers" they would want during ITS class sessions, to help them do their jobs. Overall, this exercise revealed that the analytics commonly generated by existing teacher dashboards and reporting systems for ITSs rarely align with those that teachers expect to be most useful (Holstein, McLaren, et al., 2017b). During this card sorting exercise, teachers also generated the idea of being able to see information about individual students "floating over their heads", directly within the physical classroom space. In a follow-up series of concept generation and validation studies, we found that teachers preferred *wearable* awareness tool designs (c.f., Quintana, Quintana, Madeira, & Slotta, 2016) that allowed them to keep their heads up, and their eyes focused on the classroom. Teachers emphasized that handheld real-time dashboards may compete for attention with some of the most useful real-time information in the classroom: student body language and other cues that would not be captured by a dashboard alone (Holstein, McLaren, et al., 2017b). In particular, teachers gravitated towards the idea of wearing eyeglasses that could grant them a private view of actionable information about their students, embedded through the classroom space (c.f., Alavi, Dillenbourg, & Kaplan, 2009), without revealing sensitive data to students or their peers (c.f. Jivet, Scheffel, Drachsler, & Specht, 2017).

2.2 Opening up an ITS development environment for extensible student modeling

To support the development of learning analytics tools for use with ITSs (e.g., teacher- and student-facing dashboards), my colleagues and I are substantially extending the existing CTAT and TutorShop architecture for ITS authoring and deployment (Holstein, Yu, et al., 2018c). The extended technical architecture, *CTAT/TutorShop Analytics* (CT+A), supports the authoring, sharing, and re-use of a broad and open range of student modeling techniques and analytics, for use in running ITSs (i.e., to drive adaptive tutoring behavior) and/or external learning analytics tools.

2.3 Co-designing wearable cognitive augmentation for K-12 teachers

2.3.1 Lumilo: a real-time awareness tool for personalized K-12 classrooms

Building on findings from our initial user-centered design research with K-12 teachers, briefly summarized in (2.1), my colleagues and I next conducted a series of iterative, participatory design studies with a total of 16 middle school math teachers (across 9 schools and 6 school districts in Pittsburgh and surrounding areas). We began with storyboarding, lo-fi prototyping, and participatory sketching sessions, to validate teachers' desires for real-time analytics, further probe underlying needs, and explore how teachers envisioned actually *using* this information during a class session (Holstein, Hong, Tegene, McLaren, & Aleven, 2018a). We also further explored the idea of "teacher smart glasses" further, to understand their unique affordances for monitoring personalized classes. After a Wizard-of-Oz'd mid-fidelity prototyping phase, we created a fully-functional prototype of a mixed-reality smart glasses based orchestration tool called Lumilo, capable of interfacing with a broad range of ITSs (Holstein, Hong, et al., 2018a). Using the CT+A architecture (2.2), we developed an initial set of automated detectors, using established student modeling techniques (Desmarais & Baker, 2012). To facilitate iterative prototyping of both real-time analytics and their visualizations, we also developed a novel prototyping method (discussed in the next section).

2.3.2 Replay Enactments: a prototyping method for real-time teacher support tools

To prototype the experience of using Lumilo in a classroom, we developed a new prototyping method for real-time teacher support tools: Replay Enactments (Holstein, Hong, et al., 2018a). Like other recently proposed methods in Learning Analytics (Martinez-Maldonado et al., 2016; Mavrikis et al., 2016), Replay Enactments (REs) involve replaying log data from student-software interactions, to prototype real-time analytics and their visualizations. REs go beyond these approaches by emphasizing embodied role playing and simulation exercises in physical classroom spaces, in the spirit of recent HCI methods, such as User Enactments, for prototyping radically new experiences (Odom et al., 2012). In contrast to User Enactments, however, REs prototype experiences using authentic data and algorithms, unfolding over time. Doing so facilitates observation of the interplay between teacher and machine judgments, including the UX impact of a prototype's false positives and false negatives (Dove, Halskov, Forlizzi, & Zimmerman, 2017).

2.4 Investigating relationships between teacher attention, student behavior, and student learning

To facilitate the discovery of relationships between out-of-software interactions (e.g., teacher-student interactions) and student learning within educational software in blended learning environments, we developed a new log replay tool: the Spatial Classroom Log Explorer (SPACLE) (Holstein, McLaren, & Aleven, 2017c). Using SPACLE, my colleagues and I found that students' *mere awareness* of being monitored by a teacher may contribute to student engagement and learning. We also found early evidence that, in classrooms not using a teacher awareness tool, students who exhibit patterns of "help avoidance" (Aleven, Roll, McLaren, & Koedinger, 2016) *within* educational software also tend to receive less teacher attention (Holstein, McLaren, & Aleven, 2017a).

We enhanced Lumilo to collect moment-by-moment data on a teacher's activity within the physical classroom, including gaze, position, orientation, and physical proximity to various classroom hotspots. We then conducted an in-lab experimental study, using REs, with results suggesting that Lumilo measurably directs teachers' attention towards students who would go on to exhibit lower performance on a posttest, compared with business-as-usual (Holstein, Hong, et al., 2017). Follow-up analyses using causal path modeling suggested that this effect was explained largely by Lumilo's alerts about student "unproductive persistence", or "wheel-spinning" (Kai, Almeda, Baker, Shechtman, Heffernan, & Heffernan, 2018), in educational software. To a lesser extent, this effect also appears to arise because Lumilo directs teachers' attention to students who less effectively regulate their own help-seeking behavior within the software, compared with business-as-usual (where help-avoidant students are relatively neglected) (Holstein, McLaren, et al., 2017a).

3 ONGOING WORK

3.1 A classroom experiment to study the effects of real-time teacher analytics

My colleagues and I have recently run in-vivo classroom experiments (with 286 middle school students, across 18 classrooms and 8 teachers) to investigate the effects of providing teachers with real-time analytics about student learning, metacognition, and behavior, on (1) teacher behavior; (2) student behavior and performance; and (3) students' out-of-software learning gains (Holstein, McLaren, & Aleven, 2018b). Among other findings, the results indicate that a teacher's use of *Lumilo* had a positive impact on student learning, compared with both business-as-usual and simpler classroom monitoring support. Real-time teacher analytics served as an equalizing force in the classroom: narrowing the gap in learning outcomes across students of varying prior ability.

Prior work has found that providing teachers with real-time notifications about student performance can direct their attention to *low-performing* students, resulting in local performance improvements (e.g., Martinez-Maldonado et al., 2015). Other recent work has begun systematically investigating how teachers use real-time progress and performance analytics in blended classrooms (e.g., Molenaar & Knoop-van Campen, 2017). However, the present work is the first experimental study showing that real-time teacher analytics can enhance students' *learning outcomes*.

We are currently analyzing data collected from this classroom study to better understand how the real-time analytics presented by Lumilo influenced teacher-student interactions. These classroom experiments also provided an opportunity to gather students' perspectives on the current design of Lumilo. As such, we are also currently analyzing this design feedback, to inform the design of future teacher-AI hybrid tools that can more effectively serve *students'* needs and desires.

4 PROPOSED WORK

4.1 Opening up the black box: Supporting teacher interpretation of AI inferences in teacher-AI hybrid systems.

Early in our design research with teachers, it became clear that teacher *autonomy* is a central issue in the design of teacher-AI hybrid systems (Holstein et al., 2017b; 2018a). While on the one hand teachers have often requested more direct *decision support* than is commonly offered by teacher dashboards (e.g., in the form of real-time action recommendations), especially in the face of limited time, teachers have also revealed strong discomfort with AI systems that they perceive to be "telling them what to do". During the iterative design of Lumilo, we began to explore how teacher-AI hybrid systems might effectively balance teacher autonomy with this desire for real-time decision support.

Prototyping studies with teachers (Holstein et al., 2018a) suggested that teachers' ability to *interpret* inferences and recommendations made by the AI was key not only in facilitating teacher *trust* in the AI, but also in empowering teachers to override the AI's decisions, if need be (c.f. Kamar, 2016). However, "interpretability" is a very broad notion, and in general, little is known about the effects of different forms of AI interpretability on end-users' trust, feelings of autonomy, and decision-making (Doshi-Velez & Kim, 2017; Lipton, 2016). In our prototyping studies, for example, we discovered that teachers were not particularly concerned with the *intelligibility* of an AI model (e.g., visualizations aimed at helping teachers understand *how* the AI arrived at an inference, or how the AI model was learned/trained in the first place). Instead, teachers strongly preferred *post-hoc explanations* of AI inferences (Lipton, 2016), such as curated snippets of a student's interactions with the software that could help "corroborate" a given claim made by the AI (Holstein et al., 2018a).

I propose to build on these initial investigations by systematically, empirically investigating a broader design space of mechanisms for AI to explain their own inferences (c.f. Doshi-Velez & Kim, 2017; Lipton, 2016) in teacher-AI hybrid systems, across multiple relevant evaluation criteria (e.g., effects on teacher trust in the system, as well as teachers' ability to make more informed decisions about when to override or modify AI decisions/recommendations). I expect these investigations will ultimately help pave the way for more effective and desirable partnerships between human teachers and AI systems. For example, greater interpretability may help offset otherwise harmful effects of undesirable algorithmic biases, which commonly arise in data-driven intelligent systems (Doshi-Velez & Kim, 2017; Kamar, 2016; Lipton, 2016). In addition, enabling teachers to better *understand* AI inferences may be a step towards enabling teachers to interactively provide *feedback* to these systems, to improve their usefulness within a specific classroom context (Kamar, 2016).

4.2 "Humble" AI in education: AI tutors that recognize their own limitations

The phenomenon of "unproductive persistence" (Kai et al., 2018), in AI-powered educational software can be understood as the software reaching its own pedagogical limitations (Holstein, McLaren, et al., 2017b; Käser & Gross, 2016). That is, any situation where a *learner* persists in educational software without mastering the material can also be understood as the *software* unproductively persisting in a particular teaching strategy. As such, in scenarios where a teacher is present, unproductive persistence may be viewed as a critical opportunity for the AI to (humbly) "pass control" to the teacher. Towards the design of more "humble" intelligent tutoring systems (c.f. Baker, 2016), I propose to: (1) mine pre-existing datasets, using multiple measures of "unproductive persistence" and models of other constructs (Desmarais & Baker, 2012), to better understand the *causes* of unproductive persistence within ITSs; (2) leverage findings from these analyses to develop *earlier* and *more accurate* computational methods to distinguish "unproductive" from "productive" persistence in ITSs (c.f. Kai et al., 2018); and (3) building on the work in (1), as well as section 4.1, design a system that supports teachers in interpreting the inferences made by these computational methods, and in effectively responding to instances of unproductive persistence.

REFERENCES

Alavi, H. S., Dillenbourg, P., & Kaplan, F. (2009). Distributed Awareness for Class Orchestration. In *EC-TEL* (pp. 211–225).

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help Helps, but only so Much: Research on Help Seeking with Intelligent Tutoring Systems. *IJAIED*, *26*(1), 205–223.

Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *IJAIED*, *26*(2), 600-614.

Desmarais, M. C., & Baker, R. S. J. D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *UMUAI*, *22*, 9–38.

Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *CHI* (pp. 278–288). ACM.

Doshi-Velez, F., & Kim, B. (2017). A roadmap for a rigorous science of interpretability. *arXiv preprint arXiv:1702.08608*.

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018a). Co-designing Wearable Cognitive Augmentation for K-12 Teachers. To appear in *LAK'18*. ACM.

Holstein, K., McLaren, B. M., & Aleven, V. (2017a). Informing the design of teacher awareness tools through Causal Alignment Analysis. Under Review.

Holstein, K., McLaren, B. M., & Aleven, V. (2017b). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *LAK'17* (pp. 257–266). ACM.

Holstein, K., McLaren, B. M., & Aleven, V. (2017c). SPACLE: investigating learning across virtual and physical spaces using spatial replays. In *LAK'17* (pp. 358–367). ACM.

Holstein, K., McLaren, B. M., & Aleven, V. (2018b). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. Under Review.

Holstein, K., Yu, Z., Sewall, J., Popescu, O., McLaren, B. M., & Aleven, V. (2018c). Opening up an ITS development environment for extensible student modeling. Under Review.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice. In *EC-TEL* (pp. 82-96). Springer, Cham.

Kai, S., Almeda, V.A., Baker, R.S., Shechtman, N., Heffernan, C., Heffernan, N. (2018). Modeling wheel-spinning and productive persistence in Skill Builders. To appear in *JEDM*.

Kamar, E. (2016). Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. In *IJCAI* (pp. 4070-4073).

Käser, T., & Gross, M. (2016). When to stop ? Towards Universal Instructional Policies. In *LAK'16* (pp. 289–298). ACM.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *RER*, *86*(1), 42–78.

Lipton, Z. C. (2016). The mythos of model interpretability. *arXiv preprint arXiv:1606.03490*.

Martinez-Maldonado, R., Clayphan, A., Yacef, K., & Kay, J. (2015). MTFeedback: Providing Notifications to Enhance Teacher Awareness of Small Group Work in the Classroom. *IEEE TLT*, *8*(2), 187–200.

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2016). LATUX: An iterative workflow for designing, validating and deploying learning analytics visualisations. *JLA*, *2*(3), 9–39.

Mavrikis, M., Gutierrez-Santos, S., & Poulovassilis, A. (2016). Design and Evaluation of Teacher Assistance Tools for Exploratory Learning Environments. In *LAK* (pp. 168–172). ACM.

Molenaar, I., & Knoop-van Campen, C. (2017). Teacher dashboards in practice: Usage and impact. In *EC-TEL* (pp. 125-138). Springer, Cham.

Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A., & Lee, M. K. (2012). A Fieldwork of the Future with User Enactments. In *DIS* (pp. 338–347). ACM.

Pane, J., Steiner, E., Baird, M., Hamilton, L., & Pane, J. (2017). *Informing Progress: Insights on Personalized Learning Implementation and Effects*. RAND Corporation.

Quintana, R., Quintana, C., Madeira, C., & Slotta, J. D. (2016). Keeping Watch: Exploring Wearable Technology Designs for K-12 Teachers. In *CHI* (pp. 2272–2278). ACM.

Ritter, S., Fancsali, S. E., Berman, S., & Yudelson, M. (2014). Towards Integrating Human and Automated Tutoring Systems, In *EDM* (pp. 626–627).

Rodríguez-Triana, M. J., Prieto, L. P., Vozniuk, A., Boroujeni, M. S., Schwendimann, B. A., Holzer, A., & Gillet, D. (2017). Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. *IJTEL*, *9*(2-3), 126-150.

Yacef, K. (2002). Intelligent Teaching Assistant Systems. In Kinshuk (Ed.), *IEEE* (pp. 136–140).