Informing the Design of Teacher Awareness Tools Through Causal Alignment Analysis

Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven
kjholste@cs.cmu.edu, bmclaren@cs.cmu.edu, aleven@cs.cmu.edu
Carnegie Mellon University

Abstract: Designers of teacher awareness tools (e.g., dashboards) must not only anticipate the effects analytics will have on awareness, but also how this enhanced awareness might affect teacher decision-making, and in turn, student learning. Yet teacher awareness tools are not commonly optimized to guide teacher behavior in ways that are productive for learning. In this paper, we introduce *Causal Alignment Analysis (CAA)*, a framework for data-informed, iterative design of teacher awareness tools, which links the design of awareness tools to educational goals. We illustrate the usefulness of CAA with a case study, demonstrating the successful design alignment of an awareness tool with a causal path from teacher tool use to student learning. Over a sequence of four pilot studies conducted in both simulated and live classrooms, we demonstrate the iterative refinement of *Lumilo*, a real-time awareness tool, to draw teachers' attention towards students who may benefit most from a teacher's help.

Introduction

Supporting teachers in orchestrating complex classroom activities has been identified as a key research and design challenge for the learning sciences community (STELLAR, 2011; Tissenbaum et al., 2016). In recent years, several real-time awareness tools have been designed and developed to aid teachers in orchestrating complex technology-enhanced learning scenarios (e.g., Alavi & Dillenbourg, 2012; Holstein, Hong, Tegene, McLaren, & Aleven, 2018; Martinez-Maldonado, Clayphan, Yacef, & Kay, 2015; Mavrikis, Gutierrez-Santos, & Poulovassilis, 2016). These tools augment teachers' "state awareness" during ongoing learning activities (Rodríguez-Triana et al., 2017). For example, such tools may present teachers with real-time analytics on student knowledge, progress, and metacognition within educational software (Tissenbaum et al., 2016).

The design and development of real-time teacher awareness tools is often motivated by an assumption that enhanced teacher awareness will lead to improved teaching, and ultimately, to improved student outcomes. Yet there is a paucity of evidence to support these claims, and scientific knowledge about the effects that such tools may have on teaching and learning in real educational settings is scarce (Molenaar & Knoop-van Campen, 2017; Rodríguez-Triana et al., 2017). As such, it is a challenging problem to design effective teacher awareness tools. Designers must not only anticipate the effects analytics may have on teacher awareness, but also how this enhanced awareness might affect teacher behavior, and how these changes in behavior will ultimately influence student learning. Compounding these challenges, while existing design workflows such as LATUX (Martinez-Maldonado, Pardo, Mirriahi, Yacef, Kay, & Clayphan, 2016) support the user-centered design of awareness tools based on teacher feedback, there is a lack of standard methodology for the *outcome-driven* improvement of awareness tools, to achieve targeted educational goals. Furthermore, justifications for design decisions (e.g., what information to present in a dashboard) are rarely reported in the literature (Rodríguez-Triana et al., 2017).

Researchers in other areas of educational technology research have adopted data-informed approaches to iteratively guide the design of technologies towards educational goals (e.g., Koedinger, Stamper, McLaughlin, & Nixon, 2013). For example, the design of intelligent tutoring systems (ITSs) sometimes includes an iterative refinement process, in which historical student data is leveraged to increase alignment between the software's instructional design and the way students actually learn the material, as inferred from data (e.g., Liu & Koedinger, 2017). By contrast, while teacher awareness tools are sometimes designed to be useful and usable, they are not typically optimized to guide teacher behavior in ways that are productive for learning. Given the complexity of designing teacher awareness tools, and the substantial causal distance between enhancing teacher awareness and enhancing student learning (Xhakaj, Aleven, & McLaren, 2017), bringing such outcome-driven approaches to the design of teacher awareness tools may be key to ensuring their effectiveness. In this paper, we introduce Causal Alignment Analysis (CAA): a framework for the data-informed iterative design of teacher awareness tools. We illustrate CAA via a case study, demonstrating the iterative improvement of a real-time awareness tool over a sequence of pilot studies. Finally, we discuss conclusions and directions for future work.

Causal Alignment Analysis for teacher awareness tools

Beginning from a specification of educational goals (e.g., improving student learning or engagement), CAA involves gradually aligning the design of a teacher awareness tool with these goals, by repeatedly evaluating the tool's effects along hypothesized causal paths from teacher tool use to targeted student-level outcomes. Specifically, CAA begins by generating answers to the questions below, which may represent open hypotheses where theory is absent or underspecified:

1. What student outcomes do we wish the teacher awareness tool to support?
2. What student-level processes promote or hinder progress toward the goals specified in (1)?
3. What teacher-level processes promote or hinder the student-level processes identified in (2)?
4. How can an awareness tool better support and promote the processes identified in (2) and (3)?

Taken together, answers to these questions specify hypothesized causal paths from a teacher's use of a particular awareness tool to enhanced student outcomes (as in Figure 1). Making the goals and hypothesized mechanisms of action of an awareness tool explicit early on may usefully constrain the design of an initial prototype. Once an initial prototype has been developed, Causal Alignment Analysis then involves prototyping the tool with teachers and students. Using data from these prototyping sessions, designers evaluate the alignment (or lack thereof) between the prototype's observed effects on *teacher behavior*, and one or more hypothesized causal paths to improved *student outcomes* (cf. Xhakaj et al., 2017). Based on this analysis, designers can then refine the prototype with the goal of increasing alignment, thus increasing the chances that the tool will have a positive impact in the classroom. Finally, the prototyping cycle repeats, to evaluate the effectiveness of this realignment.
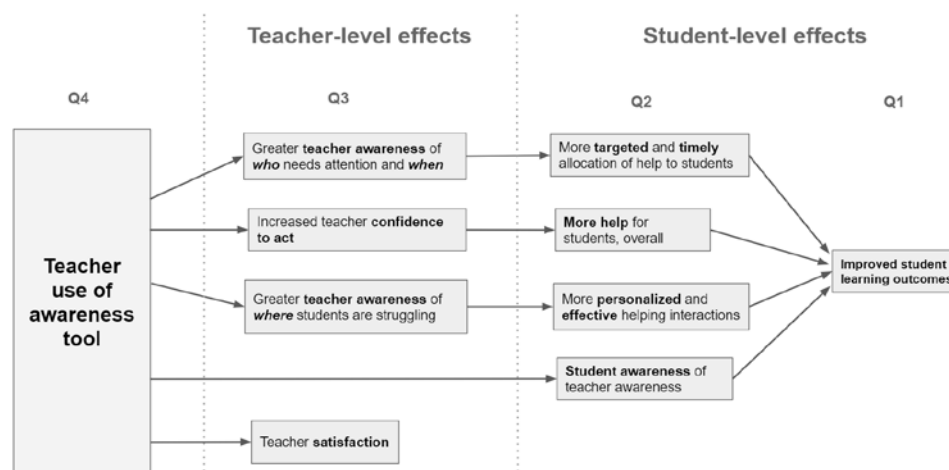

Figure 1. Examples of hypothesized causal paths, based on prior literature, leading from teacher use of an awareness tool to improved student learning outcomes. Causal tiers are labeled with questions from CAA.

Figure 1 shows examples of potential causal paths from a teacher's use of an awareness tool to teacher and student level outcomes. For these examples, we consider the context of self-paced classrooms in which students work with educational software, while a teacher uses a real-time awareness tool to decide when, with which students, and how to provide additional assistance. From left to right, the diagram shows potential influences of a teacher awareness tool (Q4) on the behavior of the teacher using it (Q3), potential impacts of resulting shifts in teacher behavior on students (Q2), and finally, potential impacts of these student-level effects on student *learning outcomes* (Q1). Given that a teacher has limited time to provide one-on-one assistance, the top path in Figure 1 posits that if teachers were alerted to critical situations (e.g., a student exhibiting a common misconception), they would be able to more effectively allocate time to students who need their attention the most, at the right moments (see Martinez-Maldonado et al., 2015). Thus, an awareness tool should be designed to alert teachers of such critical situations. In contrast to the top path – which represents a hypothesis that students using educational software would learn more from additional teacher assistance in *certain situations* – the second path, represents the hypothesis that students would benefit from more teacher attention, *in general*. Under this hypothesis, an awareness tool should be designed to encourage teachers to spend more time working with students, overall – perhaps by making teachers *feel* more informed, and thus increasing their overall "confidence to act" (van Leeuwen, Janssen, Erkens, & Brekelmans, 2015). The third causal path represents the hypothesis that, if the *quality* of a teacher's one-one-one interactions with students were improved (e.g., more tailored to a student's specific weaknesses), this would enhance student learning with the software (see van de Pol & Elbers, 2013). Furthermore, this path posits that if teachers were made more aware of student difficulties, this would lead teachers

to tailor their one-on-one interactions more closely to individual students' needs. The fourth causal path posits a *direct* link from a teacher's use of an awareness tool and a student-level effect. Under this hypothesis, students' mere awareness that a teacher is monitoring their activities in the software contributes to their learning, perhaps by increasing engagement (Holstein, McLaren, & Aleven, 2017b). Finally, the bottom path represents a hypothesis that teachers' use of a particular awareness tool positively impacts their classroom experience (Rodriguez-Triana et al., 2017), but has no notable effects on student outcomes.

Despite showing a relatively small set of hypothesized paths – each specified at a high level of abstraction – Figure 1 illustrates the enormous breadth of the design space for teacher awareness tools. Focusing on different combinations of these paths may yield radically different tool designs. In addition to guiding the initial design of a teacher awareness tool alongside user-centered design methods (Holstein et al., 2018; Martinez-Maldonado et al., 2016), CAA can be used to inform the refinement of an existing awareness tool. A designer applying CAA to the refinement of an existing awareness tool would begin by considering the tool's educational goals, and then work *backwards* from these goals (cf. Wiggins et al., 2001) to construct one or more hypothesized causal paths originating from a teacher's use of an awareness tool (guided by existing data and theory where possible). By prototyping the awareness tool, and collecting outcome data, the designer would evaluate whether the tool is likely to have desirable effects along each node in the path, adjusting the design as needed. To illustrate the use of CAA in practice, we next demonstrate the iterative design improvement of a real-time awareness tool.

Background: Co-design of a real-time teacher awareness tool

In our prior work, we designed a real-time awareness tool for teachers working in K-12 classrooms using intelligent tutoring systems (ITSs): a class of advanced learning technologies that provide students with step-by-step guidance during complex problem-solving practice. ITSs have been found, in several meta-reviews, to enhance student learning in classroom settings, compared with other learning technologies or traditional classroom instruction (e.g., Kulik & Fletcher, 2016). A key benefit of using ITSs in the classroom is that they free teachers to circulate throughout the room, providing more individualized help while students work with the software at their own pace (Schofield, Eurich-Fulcer, & Britt, 1994). However, ITSs are not typically designed to support teachers in helping their students (Holstein, McLaren, & Aleven, 2017a).

We decided to focus our awareness tool design largely on the problem of supporting teachers in allocating scarce time and attention to those students who need it the most (the top path in Figure 1), during classes in which students work individually with ITSs. This focus was motivated, in part, by user-centered design work with middle school math teachers, which highlighted these decisions as a major challenge in orchestrating personalized learning (Holstein et al., 2017a; Martinez-Maldonado, et al., 2015). In addition, this focus was motivated by prior empirical results, suggesting that teachers' decisions about whom to help, and when, may be impactful (e.g., Martinez-Maldonado et al., 2015). In particular, we focused on designing an awareness tool for classrooms using *Lynnette*, an ITS for equation solving (Long & Aleven, 2017).

In the first phase of our design process, we wanted to better understand teachers' expressed needs and desires for real-time analytics. We adopted a participatory design approach, working closely with 16 middle school math teachers (across 9 schools and 6 school districts, in a large U.S. city and surrounding areas). We directly involved teachers at each stage of the design process (cf. Martinez-Maldonado et al., 2016), including the selection and tuning of analytics through iterative user testing (Holstein et al., 2017a; 2018). The initial prototype that emerged from this iterative process was a pair of mixed-reality smart glasses (Figure 2, top-right) called *Lumilo*, which displays real-time indicators of students' current learning, metacognitive, or behavioral "processes" (as shown in Figure 2, left), floating above students' heads (Figure 2, bottom-right), while allowing teachers to keep their heads up and attention focused on the classroom (Holstein et al., 2018). The indicators displayed by the initial prototype of *Lumilo* were ideas generated and iteratively refined by teachers, and implemented using established student modeling methods (e.g., Beck & Gong, 2013; Desmarais & Baker, 2012). Together, these indicators can be taken to represent, in part, the phenomena that teachers expect require their attention and/or intervention. For example, four teachers argued that alerts about *high local error* would require immediate intervention. Otherwise, these teachers worried that repeated error-making in an ITS might entrench the errors, despite negative feedback from the software (see Metcalfe, 2017). Teachers also found some indicators valuable for other reasons. For example, we found that *positive* indicators about student performance were valuable to teachers, in part, because they found them personally motivating (Holstein et al., 2018).

Iterative improvement of Lumilo, using Causal Alignment Analysis

In addition to serving teachers' expressed needs and desires, however, we want to design awareness tools that can measurably benefit students. Teachers' intuitions about the most important opportunities for intervention may not always be correct (e.g., Baker, Walonoski, Heffernan, Roll, Corbett, & Koedinger, 2008). Therefore, in the next

phase of our design process, we used Causal Alignment Analysis to iteratively refine *Lumilo*'s design, to increase its chances of having a positive impact in the classroom. With respect to the first of CAA's four

idle
(no interactions in the software for 2 minutes or more)

rapid attempts
(many attempts or hint requests in quick succession)

help abuse or gaming the system
(e.g., systematic guessing, or misuse or unnecessary use of the software's help resources)

low local error
(low recent error rate on current activity)

high local error
(high recent error rate on current activity)

help avoidance
(not using hints, when they are likely needed)

high error after hints
(high error rate, even after using all hints on a given step)

unproductive persistence
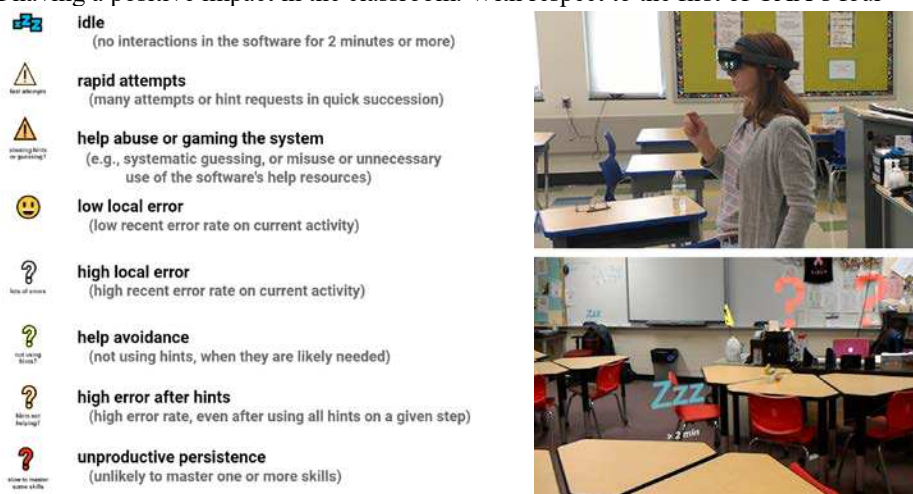(unlikely to master one or more skills)

Figure 2. Left: Full set of student-level indicators displayed by an early version of *Lumilo*. Top-right: Teacher using *Lumilo*. Bottom-right: Point-of-view screenshot (moments after the end of class).

guiding questions, we had defined our learning objectives as the set of equation-solving skills that *Lynnette* tutors. In answer to CAA's second and third questions, we adopted a causal model search approach to understand the relationships between *Lumilo*'s indicators and student learning outcomes – hypothesizing that teacher attention should be directed to student processes with a negative influence on learning. Finally, in response to CAA's fourth question, we iteratively refined *Lumilo* to direct teachers' time and attention towards these processes, over a sequence of in-lab and classroom pilot studies. Each step is discussed next, in turn.

To answer CAA's second question ("What student processes promote or hinder ..."), we sought to better understand the relationships between student processes detected by the current prototype of *Lumilo* (the student-level indicators shown in Figure 2, emerging from our participatory design process) and student *learning* within *Lynnette*. To this end, we adopted a causal model search approach, using directed acyclic graphs (DAGs) to represent the causal structure among variables measured by *Lumilo*, and student assessment scores. We collected data from 115 middle school math students (across 7 classrooms and 4 teachers), each of whom worked with *Lynnette* for 60 minutes. In these classrooms, the teacher did not use an awareness tool (Table 1, Study 1). In all studies, we assessed students' equation-solving skill with a pretest and posttest administered before and after using the tutor. We used two forms that were identical except for the specific numbers used in equations. We presented the forms in counterbalanced order across pre- and posttest.

We then used the PC algorithm in the Tetrad V program to search for an equivalence class of DAGs, consistent with a set of conditional independence constraints (Spirtes et al., 2000). The PC algorithm is asymptotically reliable; its primary limitations are its assumptions that no unmeasured confounders are present, and that any underlying causal relationships between variables can be modeled by linear functions. To relax the former of these assumptions, we also used the FCI algorithm, which allows for the possibility of unmeasured confounders. The FCI algorithm learns an equivalence class, represented by partial ancestral graphs (PAGs), encoding uncertainty over the nature of pairwise relationships between variables (Spirtes et al., 2000). To inform both searches, we provided background knowledge about our study design as a search constraint: we specified that the pretest was prior to any process variables, and that all process variables preceded the posttest.

Figure 3 (left) shows the DAG learned with the PC algorithm, including normalized coefficient estimates, to enable comparison of magnitudes. This model suggests that, of the indicators included in the initial prototype of *Lumilo*, three are potential direct causes of reduced student learning within the software: *help abuse or gaming-the-system* (measured by the Help Model and gaming detector, reviewed in Desmarais et al., 2012), *high local error* (defined by teachers as an error rate greater than 80%, within the last 8 student actions on the current activity), and *unproductive persistence* (measured by the "wheel-spinning" detector, described in Beck & Gong, 2013). This model fits the data well ($\chi 2 = 18.33$, $df = 19$, $p = .50$) (1). Figure 3 (right) shows the PAG learned with the FCI algorithm. In this figure, bidirectional links indicate the presence (and circle-origin links indicate the *possibility*) of unmeasured confounders. Otherwise, links indicate causal relationships. Wide links indicate no unmeasured confounders, and dark, wide links further indicate *direct* relationships. The PAG equivalence class found by FCI suggests that unmeasured confounders could potentially explain several of the links between

Lumilo's indicators. Finally, in both causal models, *gaming/help-abuse*, *unproductive persistence*, and *help avoidance* (as measured by the Help Model, Aleven et al., 2016) are negatively linked to
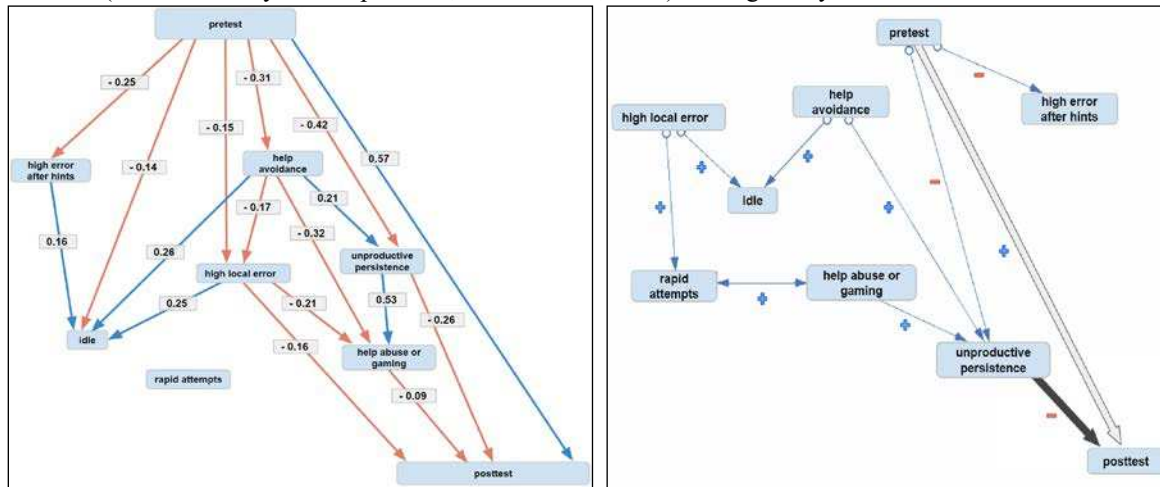

Figure 3. Left: model found by PC, with normalized coefficient estimates included. Right: PAG equivalence class found by FCI, encoding the possibility of unmeasured common causes.

student learning. The model found by FCI suggests that out of 7 *negative* teacher-generated indicator ideas implemented in *Lumilo*, only one is directly linked to student learning: *unproductive persistence*. Influences of help avoidance and gaming/help-abuse on learning may in turn be mediated through unproductive persistence.

To determine how the design of *Lumilo* might be improved (the fourth question in CAA), we wanted to first understand how the *current* prototype of *Lumilo* influences teacher behavior, prior to deploying it in real classrooms. To this end, we conducted a series of simulated class sessions using a new prototyping method called Replay Enactments (REs) (Holstein et al., 2018). In each session, historical student interaction data were replayed in ITS interfaces, on separate computer screens in a classroom setting (but with no actual students present). Following a 35-minute training period in which teachers acclimated to using the tool and studied the definitions of each of *Lumilo*'s indicators, each teacher participated in a 40-minute replay session. In these sessions, teachers wore *Lumilo*, and were asked to think aloud while monitoring the "class". If a teacher thought they would intervene with a certain "student" at a given time, the teacher would approach that "student" and enact the help session aloud. In addition to recording think-aloud data, we used *Lumilo* to automatically track the teacher's physical position moment-by-moment (Holstein et al., 2018).

First, we investigated how teachers' time allocation across students during REs may have been influenced by each of *Lumilo*'s student-level indicators. Teacher time allocation was measured per student by the cumulative time (in seconds) spent within a 4-ft. radius of that student (resolving ties among students by proximity), as well as time spent monitoring the student's activities via *Lumilo*'s deep-dive screens (Holstein et al., 2018). Table 1 (Study 2) shows group-normalized correlations between detected student processes and teachers' time allocation during six REs. Real-time indicators that were not significant predictors of teacher time allocation are omitted. As shown, occurrences of four of *Lumilo*'s indicator alerts were significantly positively correlated with teacher time allocation. Second, to understand the degree to which the awareness tool might have directed teachers towards students most in need of help, as per the top path in Figure 1, correlations between student assessment scores and teacher time allocation are also shown in Table 1. Given that teachers did not have access to assessment scores during REs, and that it is not possible to influence learning during a *replayed* class, we take the correlation between teacher time allocation during REs and student posttest scores (controlling for pretest) as evidence that *Lumilo* can direct teachers' time to students who would otherwise exhibit lower learning. However, this correlation was relatively small, suggesting room for improvement.

Taken together, these analyses suggested various ways the design of *Lumilo* could be improved (Q4), to increase its alignment with the hypothesized causal path shown in Figure 4. Unproductive persistence was the weakest driver of teacher attention during REs, out of the indicators correlated with teacher time allocation (as shown in Study 2 of Table 1), despite being the one variable directly (and negatively) related to student learning in the causal model found by FCI. To better align *Lumilo*'s design with these analyses, the design should focus more explicitly on alerting teachers to cases of unproductive persistence, by increasing the salience of this alert and others that may serve as reliable early predictors. For instance, although help avoidance is a potential cause of unproductive persistence in the PAG found by FCI (and thus potentially valuable as an early predictor), it was

not a significant driver of teacher attention. Similarly, this model suggests that less emphasis should be placed on alerting teachers to *high local error* or *rapid attempts* in general, and more should be placed on alerting teachers to cases that constitute maladaptive help-use and/or gaming (Desmarais & Baker, 2012). As such, we next refined the prototype of *Lumilo* to place greater emphasis on alerts about unproductive persistence and persistent help avoidance. This included not only making the corresponding indicator symbols more visually salient than others (larger and brighter), but also drawing teachers' attention to these alerts through ambient sound notifications. Meanwhile, we de-emphasized other alerts, including *high local error* and *rapid attempts* by making these indicators relatively dimmer and smaller. Furthermore, if a student was detected as unproductively persisting on one or more skills, avoiding help, or gaming/abusing-help, any other alerts for that student would be *hidden* at a glance (although still accessible upon a teacher's request).

Table 1: Correlations between teacher time allocation, and detected student processes and test scores, * *p* < 0.05, ** *p* < 0.01, *** *p* <0.001. Rows show a series of studies, using successive versions of *Lumilo*

	Study Context				Process Variables (awareness tool alerts)					Assessment Scores	
Stu-dy	type	aware-ness support	sample (teachers, classes, students)	total time (min)	unpro-ductive persist-ence	help avoid-ance	help abuse or gaming	rapid attempts	high local error	pretest	post | pre
1	live	none	(4, 7, 115)	60	- 0.06	**- 0.17***	- 0.09	- 0.12	0.06	0.13	- 0.02
2	RE	*Lumilo v1*	(6, 3, 90)	40	**0.25***	- 0.03	**0.44*****	**0.38*****	**0.32****	**- 0.06****	**- 0.17****
3	live	*Lumilo v2*	(1, 1, 15)	40	**0.65***	**0.61***	0.22	0.27	0.39	**- 0.84*****	0.40
4	live	*Lumilo v3*	(2, 4, 84)	60	**0.52*****	**0.16***	**0.07***	0.01	0.18	**- 0.30***	0.16

We next ran two more pilot studies, in live classrooms. The first of these studies was run with one teacher in a single, 80-minute class session. In this study, students worked with *Lynnette* for 40 minutes, while the teacher used *Lumilo* (version 2) to monitor and help students. Students' domain knowledge in equation solving was measured before and after using the software, via computer-based pre- and posttests, as in prior studies. As shown in Study 3 of Table 1, students who were more frequently detected as unproductively persisting or avoiding help received significantly more teacher time during this single-classroom pilot, compared with students exhibiting other behaviors tracked by *Lumilo*, suggesting that the design refinements may have had the intended effect. Furthermore, the teacher's attention during this single-classroom pilot was strongly and significantly focused towards students with lower prior domain knowledge (as measured by the pretest), and the correlation between teacher time allocation and student posttest score (controlling for pretest) was positive, despite a likely selection effect, although not statistically significant.

Following this pilot, we made minimal design refinements to *Lumilo*, in an effort to ensure that alerts of unproductive persistence were emphasized (as potentially more critical) over alerts of help avoidance and gaming/help-abuse. In version 3 of *Lumilo*, if a student was detected as unproductively persisting in the software on one or more skills, any other alerts for that student would be hidden. We ran additional classroom pilots using *Lumilo* (version 3) in 4 classrooms. Students in each classroom worked with *Lynnette* for a total of 60 minutes while the teacher used *Lumilo* to monitor and help their students. As before, student domain knowledge was measured via 20-minute, computer-based pre- and posttests. As shown in Study 4 of Table 1, *unproductive persistence* was the strongest predictor of teacher time allocation, followed by *help avoidance* and *gaming/help-abuse*. Classroom observations indicate that teachers continued to make use of all indicators presented by *Lumilo* (e.g., praising recent high performers or nudging inactive students), but tended to reserve in-depth remediation sessions for those students detected as unproductively persisting. Retrospective post-interviews corroborated this observation. However, teachers also reported frequently attending to "quick fix" alerts for students physically "en-route" to a particular student the teacher was targeting for remediation.

In summary, in the first phase of our design process, we decided to focus on the problem of supporting teachers in allocating scarce time and attention to those students who may need it most. We adopted a participatory design approach, eliciting ideas for real-time analytics that teachers considered actionable, relevant to learning, or otherwise valuable to monitor. We leveraged pre-existing student modeling techniques to provide teachers with

these analytics, while iteratively prototyping them with teachers to ensure their usefulness and usability. In the next phase of our design process, we used CAA to iteratively align *Lumilo*'s design with a hyp-
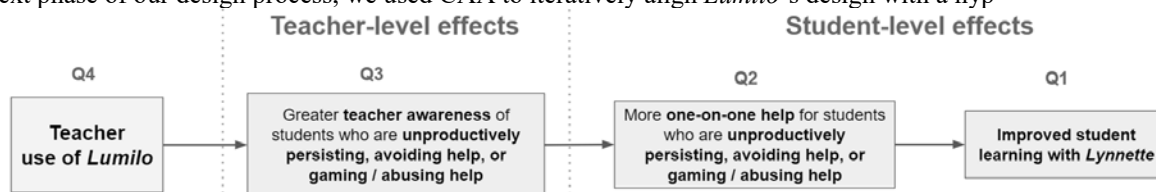

Figure 4. Hypothesized causal path from a teacher's use of *Lumilo* to improved student learning.

othesized causal path to improved learning outcomes, learned from data (a finer-grained instantiation of the top path in Figure 1, as shown in Figure 4). With respect to the first of CAA's four guiding questions, we defined students' learning objectives as the skills that *Lynnette* is intended to tutor, and assessed student learning with respect to these skills. In answer to CAA's second and third questions, we adopted a causal model search approach to discover a critical subset of *Lumilo*'s indicators, representing student processes that most strongly influence learning outcomes with *Lynnette*. In turn, we hypothesized that students exhibiting these processes may benefit most from out-of-software, teacher interventions. Finally, with respect to CAA's fourth question, we iteratively refined *Lumilo* – over a sequence of four pilot studies conducted in both simulated and live classrooms – to draw teachers' time and attention towards these students.

Conclusions and future work

In this paper, we have introduced Causal Alignment Analysis (CAA): a design framework for the data-informed design and iterative improvement of teacher awareness tools, linking the design of these tools to educational goals. We have illustrated the application and usefulness of CAA through a case study, demonstrating the iterative design alignment of a real-time teacher awareness tool with a hypothesized causal path from teacher tool use to student learning (Figure 4). The resulting prototype augments teachers' awareness of student learning, metacognition, and behavior, while also measurably directing their time towards a subset of student processes that appear to have a negative influence on student learning outcomes.

While this case study may represent a step towards the design of teacher awareness tools that can measurably enhance student learning, it does not fully "close the loop" (Koedinger et al., 2013). To support iterative design, a CAA approach favors running larger numbers of small to mid-scale studies over running a single high-powered study. As such, it may not support strong causal inference. To better understand whether and how a teacher's use of *Lumilo* influences student learning, we have recently conducted a larger-scale classroom experiment. Analyses of data from this experiment will enable us to investigate multiple hypothesized paths from teacher tool use to student learning (Figure 1), and thus to tease apart the distinct causal explanations that these paths represent. For example, although the analyses presented in this paper led to the improvement of *Lumilo* with respect to the hypothesized causal path pictured in Figure 4, it remains an open question whether the final link in this path (improved student learning) will hold in practice.

While the case study presented in this paper focused on data-informed design optimization with respect to teacher attention allocation across students (the top path in Figure 1), there are many other causal paths along which an awareness tool might be optimized. For instance, even if teachers are made more *aware* of critical moments, it may not always be clear how to effectively *respond*. Our design work with teachers suggests that they often desire more direct support (e.g., action recommendations) for planning and enacting effective interventions – especially in personalized learning contexts, where planning time can be very scarce (Holstein et al., 2018). A promising direction for future work may be to use CAA to explore whether and how an awareness tool could be designed to measurably enhance the *effectiveness* of teacher-student coaching interactions.

In summary, as the fast-growing research area of teacher awareness tools matures, we hope to see the design of these tools (within and beyond the academic Learning Sciences and Learning Analytics communities) increasingly guided by educational data and theory, *in addition to* user feedback. Causal Alignment Analysis provides a framework for making the goals and implicit assumptions behind the design of awareness tools explicit – in turn representing these assumptions as hypotheses to be continuously tested throughout a design process. Given the complexity of designing teacher awareness tools, we expect that such data-informed design approaches will be key to ensuring that they are not only useful and usable, but also beneficial for learning.

Endnotes

(1) In path analysis, the null hypothesis is that the estimated model is the true model. The p-value represents the probability, under the null, of observing a difference between the estimated and observed covariance matrices at least as large as the realized difference; a p-value above a given threshold (conventionally alpha = .05) implies a model *cannot* be rejected.

References

Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE TLT*, *5*(3), pp. 264-274.

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), pp. 205–223.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *JILR*, *19*(2), pp. 185-224.

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. In *AIED* (pp. 431–440). Springer, Berlin, Heidelberg.

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*, pp. 9–38.

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. To appear in *LAK'18*. ACM.

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

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

Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using data-driven discovery of better student models to improve student learning. In *AIED* (pp. 421-430). Springer, Berlin, Heidelberg.

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

Liu, R., & Koedinger, K. R. (2017). Closing the Loop: Automated Data-Driven Cognitive Model Discoveries Lead to Improved Instruction and Learning Gains. *JEDM*, *9*(1), pp. 25-41.

Long, Y., & Aleven, V. (2017). Educational Game and Intelligent Tutoring System: A Classroom Study and Comparative Design Analysis. *ACM TOCHI*, *24*(3), 20.

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), pp. 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), pp. 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.

Metcalfe, J. (2017). Learning from Errors. *ARP*, *68*(1), pp. 465–489.

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

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), pp. 126–150.

Schofield, J. W., Eurich-Fulcer, R., & Britt, C. L. (1994). Teachers, Computer Tutors, and Teaching: The Artificially Intelligent Tutor as an Agent for Classroom Change. *AERJ*, *31*(3), pp. 579–607.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. MIT press.

STELLAR. (2011). Trends in orchestration. Second research & technology scouting report.

Tissenbaum, M., Matuk, C., Berland, M., Lyons, L., Cocco, F., Linn, M., …, Dillenbourg, P. (2016). Real-time visualization of student activities to support classroom orchestration. In *ICLS 2016* (pp. 1120–1127).

van de Pol, J., & Elbers, E. (2013). Scaffolding student learning: A micro-analysis of teacher–student interaction. *LCSI*, *2*(1), pp. 32–41.

van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *C&E*, *90*, pp. 80–94.

Wiggins, G., & McTighe, J. (2001). What is backward design?. In *Understanding by Design* (pp. 7–19).

Xhakaj, F., Aleven, V., & McLaren, B. M. (2017). Effects of a Teacher Dashboard for an Intelligent Tutoring System on Teacher Knowledge, Lesson Planning, Lessons and Student Learning. In *EC-TEL* (pp. 315–329). Springer, Cham.

Acknowledgments