Intelligent Tutors as Teachers' Aides: Exploring Teacher Needs for Real-time Analytics in Blended Classrooms

Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{kjholste, bmclaren, aleven}@cs.cmu.edu

ABSTRACT

Intelligent tutoring systems (ITSs) are commonly designed to enhance student learning. However, they are not typically designed to meet the needs of teachers who use them in their classrooms. ITSs generate a wealth of analytics about student learning and behavior, opening a rich design space for real-time teacher support tools such as dashboards. Whereas real-time dashboards for teachers have become popular with many learning technologies, we are not aware of projects that have designed dashboards for ITSs based on a broad investigation of teachers' *needs*. We conducted design interviews with ten middle school math teachers to explore their needs for on-the-spot support during blended class sessions, as a first step in a user-centered design process of a real-time dashboard. Based on multi-methods analyses of this interview data, we identify several opportunities for ITSs to better support teachers' needs, noting that the analytics commonly generated by existing teacher support tools do not strongly align with the analytics teachers expect to be most useful. We highlight key tensions and tradeoffs in the design of such real-time supports for teachers, as revealed by "Speed Dating" possible futures with teachers. This paper has implications for our ongoing co-design of a real-time dashboard for ITSs, as well as broader implications for the design of ITSs that can effectively collaborate with teachers in classroom settings.

CCS Concepts

- **Human-centered computing ~ User centered design**
- **Applied computing ~ Computer-assisted instruction**

Keywords

Intelligent tutoring systems, pedagogical decision-making, real-time analytics, blended learning, teachers, classrooms, adoption

1. INTRODUCTION

In recent years, there has been growing interest in *teaching analytics*: the use of analytics to support teacher awareness, reflection, and decision-making in both physical and virtual classroom settings [2, 21, 22, 24]. In particular, there has been an increasing interest in the design and development of systems that can support teachers' on-the-spot decision-making, by presenting them with actionable analytics in real-time (e.g., [2, 13]). Some of this work has focused specifically on supporting teacher monitoring and decision-making in blended learning environments, where students may work with adaptive learning technologies at their own pace. A key advantage of such classroom technologies is that they free the teacher to provide more one-on-one support to students who may benefit from it the most. However, they also present teachers with unique challenges, as teachers are tasked with monitoring classrooms that may be working on a broad range of divergent educational activities simultaneously, and prioritizing help-giving across students, in the face of limited time [2, 27].

We are working towards the design of a real-time dashboard for teachers in K-12 classrooms who use adaptive educational technologies [15, 19] as part of their instruction. In particular, we are exploring how intelligent tutoring systems (ITSs) might be better designed to meet the needs of teachers during blended class sessions, and how a real-time dashboard might meet some of these needs. ITSs are a type of adaptive educational technology that provide students with detailed, step-by-step feedback during complex problem-solving practice, while adapting instruction based on continuously-updated models of students' current state. These student models may include moment-by-moment estimates of student knowledge (e.g., [38]), whether a student seems "stuck" on a given skill (e.g., [19, 32]) or engaged in their current activity (e.g., [34]), whether a student is exhibiting particular misconceptions (e.g., [3]), and so on.

Despite several meta-reviews indicating that ITSs can significantly enhance student learning, compared with traditional classroom instruction and other types of educational technologies, these systems have thus far struggled to achieve high adoption [4, 15] (though see [20, 35, 36]). A recent systematic review of the ITS literature, conducted by Nye, assessed the amount of attention the field has paid to various potential barriers to broader ITS adoption. In this work, Nye considered barriers at multiple levels relevant to technology adoption and use (student, teacher, and school), and highlighted the relative rarity of research on monitoring and customization tools designed to help teachers integrate ITSs into their pedagogy. Noting that systems with higher adoption (e.g., Cognitive Tutor, ASSISTments, and ALEKS) tend to have some form of either monitoring or customization tools for teachers, Nye speculated that under-consideration of teachers' needs in classrooms using ITSs may be a significant factor affecting adoption and attrition [4]. Similarly, other authors have speculated that a key barrier to adoption may be that ITSs have not typically been designed with a focus on teachers' needs in blended classrooms [3, 9, 10, 26].

LAK '17, March 13 - 17, 2017, Vancouver, BC, Canada
ACM 978-1-4503-4870-6/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3027385.3027451

Over a decade ago, Yacef proposed a reframing of intelligent tutoring systems as “intelligent teaching assistants” (ITAs): intelligent systems that are designed with the dual objectives of helping human teachers teach and helping students learn (rather than just the latter objective, as is typical of ITSs) [7]. Others have proposed similar directions, from the lens of optimizing student learning: ITSs might better facilitate student learning if these systems could both help and *leverage* human teachers, when teachers are present in the classroom [4, 9, 11, 28]. While there has been some work on real-time teacher support tools for ITSs since the vision of ITAs was introduced (e.g., [8, 25, 29]), the choices of analytics presented within these tools often appear to have been influenced more by the availability of data, rather than by an analysis of teachers’ needs (c.f. [24]). Recent work has explored the use of participatory design approaches to inform the design of real-time teacher support tools for blended learning environments (e.g. [30]). However, while this work focused on understanding how teachers would use an *existing* prototype of a real-time dashboard to inform their pedagogy, in the present work we aim to understand teachers’ needs for real-time support tools in classrooms using ITSs *prior* to designing any such tools.

The structure of this paper is as follows: to better understand how current ITSs may meet some of teachers’ needs, while failing to meet others, we first present a brief case study of technology-use practices and breakdowns among a group of middle school math teachers who, for five years, used a widely-used ITS as a core component of their regular classroom instruction. We share findings from semi-structured interviews with two of these teachers, examining the evolution of their classroom work practices over this five-year period, as well as the reasons behind their decision to discontinue use of the software and its associated curriculum. As a first step in our user-centered design of a real-time teacher’s dashboard, we then present a series of design exercises that we conducted with teachers in order to better understand their needs in classrooms that use ITSs. We discuss the implications of our findings for the design of ITSs that can facilitate synergistic interactions between K-12 teachers and their students, via real-time analytics. In order to explore potential barriers to adoption of such real-time supports, and also to detect unexpected design opportunities, we then “speed date” a number of imagined futures with teachers (presented as concept sketches and storyboards) [5, 6]. Finally, we briefly discuss the broader implications of these findings for the design of intelligent tutoring systems that can effectively collaborate with human teachers, and present directions for future work.

2. CASE STUDY: A FIVE YEAR RELATIONSHIP BETWEEN TEACHERS AND INTELLIGENT TUTORS

To gain a better sense of the ways current ITSs may meet some of the needs of K-12 teachers but fail to meet others, we interviewed two mathematics teachers (teachers 8 and 9 in Table 2) from a school in the Pittsburgh area (school E in Table 1), who had used an ITS in their classrooms for about five years. The interviews were semi-structured, and incorporated a version of the *Love Letter and Breakup Letter* design method, which uses personification as a tool to probe participants’ original reasons for adopting a technology (and continuing to use it for an extended period) and their reasons for eventually “breaking up” with that technology [6]. Our findings from these interviews are briefly summarized below, charting these teachers’ journey from adoption, to break-up, and then to the present.

According to the interviewed teachers, teachers at this school originally pushed to adopt this ITS and its associated curriculum as part of a larger teacher-led effort to move away from their existing mathematics curriculum. These teachers felt that the existing curriculum involved too much interleaving of topics, in which “you never really get fully though a topic the first time, or the second time, or the third time…”. Instead, they wanted to move to a curriculum that “[teaches] a topic *once*, [making sure that] they master it”, and then allows students and teachers to move on. They found this mastery learning approach to mathematics instruction appealing largely because they felt it represented the sort of deeper, more focused learning that students would be required to do in high school and beyond. As such, they decided to adopt the ITS – which implemented a mastery learning approach to activity sequencing – in part to provide their students with extra scaffolding during the transition from shallower and less independent learning (in elementary school) to deeper and more self-regulated learning (in high school).

Table 1. Demographic information for schools

School	Region	Free/Reduced Price Lunch
A	Suburban	17.7%
B	Urban	n/a
C	Suburban	23.4%
D	Suburban	29.3%
E	Rural	30.7%

The interviewed teachers noted that the first year of using the ITS in their classrooms was a challenging adjustment period. Despite the support materials that came with the ITS at that time, including a curriculum with associated materials such as textbooks, professional development, and a reporting system that allowed teachers to track their students’ progress regularly, these teachers initially struggled to decide how best to monitor and help their students during lab sessions in which students used the ITS. In particular, they had trouble determining how to assess students fairly and accurately given the self-paced nature of adaptive learning technologies. A major constraint teachers face (in most US K-12 schools, at least) is that they need to provide students with letter grades and to communicate their reasons for assigning a particular grade to both students and their parents. During this first year, these teachers often found that it was difficult to *justify* their decisions to assign students grades based on their progress within the ITS – particularly when communicating with these students’ parents. Teachers’ grading decisions often involved a considerable amount of subjectivity, as it was often unclear how to balance between grading students based on the *progress* they’ve made in the software (i.e. how many units of the curriculum a student has covered), how well students have *performed* within those units (as shown in the software reports as probabilities that a student has mastered each of a number of fine-grained skills), and how much *growth* students have shown individually (i.e., change in students’ per-skill probabilities of mastery over time).

After the first year of use, teachers began to hold meetings to share reflections, insights, and strategies on how to most effectively use the ITS in their classrooms. Through these meetings, teachers at this school collectively developed common work practices and grading procedures to help mitigate some of the major challenges they had encountered over the previous year. For example, these teachers developed a uniform grading scheme by setting goals for where students should be (in terms of the number of units covered) at regular checkpoints throughout each semester. The teachers decided upon these goals by pooling their recollections of the unit most students had reached in the previous year, by certain checkpoints (e.g., the beginning of each month). If a student was one or more units behind the goal unit, at a certain checkpoint, the teacher would use a control panel in the software to manually push that student forward to the goal unit (and the student would receive no credit for any intervening units). This was done to keep the whole class relatively synchronized over the course of the year, and to manage the complexity of assigning individual letter grades when different students have covered different amounts of material. Over time, teachers' grading schemes became more nuanced, as teachers would annotate printed versions of the ITS-generated reports with their own observations, collected each day while monitoring their classrooms. They would sometimes integrate these annotations with the ITS-generated metrics to assign grades – allowing for partial credit to be given based on their perception of the student's effort or students' *growth* over time, rather than just the speed at which they reached mastery.

According to the interviewed teachers, they (and other teachers at the school) ultimately agreed that continued use of the ITS was not worth the cost, for three primary reasons:

1. Challenges of curriculum alignment. Late in the five-year use period, the school district began a shift to a new mathematics curriculum, and the teachers needed to drop the curriculum that came with the ITS. During this time, teachers increasingly found that it was challenging to align the school's new mathematics curriculum with the content and instructional design of the ITS. Yet there was no convenient way for teachers to customize the ITS's content to meet their changing needs. One teacher suggested that the ability to make relatively minor customizations to the ITS's problem interfaces (e.g., editing the way math problems were represented, and altering the input format that the ITS would accept from students) would have helped, but only if such customizations could be made very quickly.

2. Semi-manual grading and monitoring systems were difficult to maintain. Although the ITS generated detailed reports about students' progress and performance within the software (e.g. probabilities that a student had mastered finely-defined skills in the curriculum, and the number of hints a student had requested), teachers noted that these reports did not provide them with guidance about how to fairly and accurately assign students letter grades based on the data. As such, the teachers felt the need to develop their own grading system, which necessarily balanced efforts to be fair and accurate with teachers' time constraints. Another key limitation these teachers highlighted was that it was not always easy to identify students who were falling behind until it was already "too late" for the student to catch up with the rest of the class. That is, the most salient elements of the reports provided by the ITS tended to be information about the past (e.g. that a student had been overusing the ITS's hints, or that a student had *not yet* mastered finely-defined skills in the curriculum). But these reports typically did not provide *predictive* analytics that could help teachers anticipate problems and proactively intervene. One teacher noted that they would have liked to be able to see the likelihood that a student who had fallen behind the class would actually be able to "catch up" with the other students, if given more time. Without this information, pushing a student forward almost always seemed like the most reasonable decision.

3. Perceived susceptibility of these systems to student misuse. Some of the teachers perceived that ITSs are particularly susceptible to "gaming" or "cheating" (e.g. abusing the hints that the ITS provides, or solving a math problem via a brute force approach). These teachers believed that, since they had often been unable to catch these behaviors in a timely manner, some of their students had likely wasted a large amount of learning time. Research supports these teachers' intuitions to a degree: gaming behaviors in ITSs have consistently been shown to have a negative impact on student learning, overall (although not all gaming behaviors are necessarily harmful, and only a relatively small proportion of students has a tendency to game the system) [37]. These teachers were also skeptical that a fully automated mechanism could prevent students from gaming. One of the interviewed teachers suggested that alerts about such misbehavior, which are easily hidden in large computer labs, should be sent to the teacher right away.

Reasons 1 and 2 correspond closely with two of the issues that Nye highlighted as relatively under-considered in the ITS literature – namely the lack of sufficient customization and monitoring capabilities [4]. Each of these cases can be viewed as an instance of the teacher adapting to the technology, rather than the other way around [3, 26]. The length and difficulty of teachers' adjustment to the use of ITSs in their instruction may also highlight a need for enhanced early support, in the form of improved teacher training tools and peer support systems that facilitate faster sharing of strategies and observations between teachers (as teachers eventually felt the need to band together, but did so only after significant struggle). Teachers' practice of "pushing students forward" when they do not achieve mastery quickly enough represents an interesting case, as recent research suggests that such teacher "overrides" of ITSs' mastery learning algorithms can be quite harmful to student learning over the course of a school year [31]. This points both to a need for caution in designing such customization and control options for teachers, as well as a need to better understand the constraints and beliefs that might lead teachers to make such decisions. Although teachers were aware that the practice of pushing students forward before they had mastered the skills in a given unit was counter to the idea behind mastery learning, they continued to do so in order to keep the class relatively synchronized and manage their own orchestration load.

The interviewed teachers also noted that, since discontinuing use of the ITS, they had not adopted any other learning technologies in their classrooms. They emphasized that they had used the system for many years because they believed the personalized, detailed, and immediate feedback it provided to students was very valuable for their learning. In fact, they strongly preferred using ITSs to other educational technologies they had tried over the years, for this reason. The primary obstacles to teachers' continued use of these systems did not lie in the perceived effectiveness of ITSs, but rather in the difficulties that their use in the classroom presented for teachers.

3. INVESTIGATING TEACHERS' WANTS AND NEEDS

3.1 Methods

To gain a better sense of teachers' needs in blended classrooms using ITSs, we conducted four types of design interviews (generative card sorting exercises, semi-structured interviews, directed storytelling, and Speed Dating sessions), with a total of 10 middle school teachers, from five schools in Pittsburgh and surrounding areas. All of the participating teachers had previously used at least one adaptive educational technology in their classrooms, and nine out of ten teachers had previously used an ITS at least once (though only teachers from schools D and E had used ITSs as a regular component of their teaching).

Table 2. Demographic information for teachers

Teacher	Gender	Teaching experience (years)	School
1	Male	> 20	A
2	Female	2	B
3	Male	> 20	C
4	Female	> 10	C
5	Male	2	E
6	Male	25	E
7	Female	11	D
8	Male	16	D
9	Male	15	D
10	Female	19	D

3.2 "Superpowers" as a probe to investigate teachers' major perceived challenges

We wished to start by investigating teachers' needs for real-time support during blended class sessions, in a very broad sense. To this end, we adopted a card generation and sorting approach [6, 14]. In separate sessions, we met with five teachers from two schools (teachers 4, 7, 8, 9, and 10, from schools C and D) and asked them, "If you could have any superpowers you wanted, to help you do your job, what would they be?". We asked the question this way to encourage teachers to talk freely about their needs, and breakdowns in their current practices, without feeling constrained to those for which they believed a technological solution was possible. Although we initially asked this question in a very broad sense, we gradually narrowed the question's focus to superpowers that teachers would find useful specifically during computer lab sessions in which students are interacting with an ITS or other adaptive educational technology.

In each interview, we asked middle school teachers to write their desired superpowers on index cards immediately upon generating them (to reduce the chance that they would lose track of an idea as the conversation progressed). In addition to identifying design opportunities within the sets of cards teachers generated, we wished to get a better sense of teachers' *priorities* among superpowers (and by proxy, the relative severity of the daily challenges behind these "superpower requests"). To this end, once a teacher had finished generating ideas for superpowers, they were asked to sort them by subjective priority, from highest to lowest [6, 14]. Throughout the card sorting process, teachers were encouraged to generate new cards, if this process inspired new ideas. Then, following this initial sorting, each teacher was presented with cards generated by all previous teachers who had participated, and was given the option to include any of these in their hierarchy. If a teacher felt that one of these superpowers was redundant with one of the superpowers they had generated, they were encouraged to align these cards horizontally, to indicate a tie. For any superpowers a teacher did not desire, the teacher was asked to omit that card from the hierarchy. Figure 1 shows an example of a hierarchy resulting from this iterative card generation and sorting process. This teacher's most desired superpower was the ability to "put the info directly into [students'] heads", without any need to interact with students. Interestingly, no other teacher included this particular superpower. One teacher defended this omission by suggesting that having such a power would remove the joy from teaching. This teacher's second most desired superpower was omniscience, which the teacher considered synonymous with the ability to "see students' thought processes" (a card generated by a previous teacher).

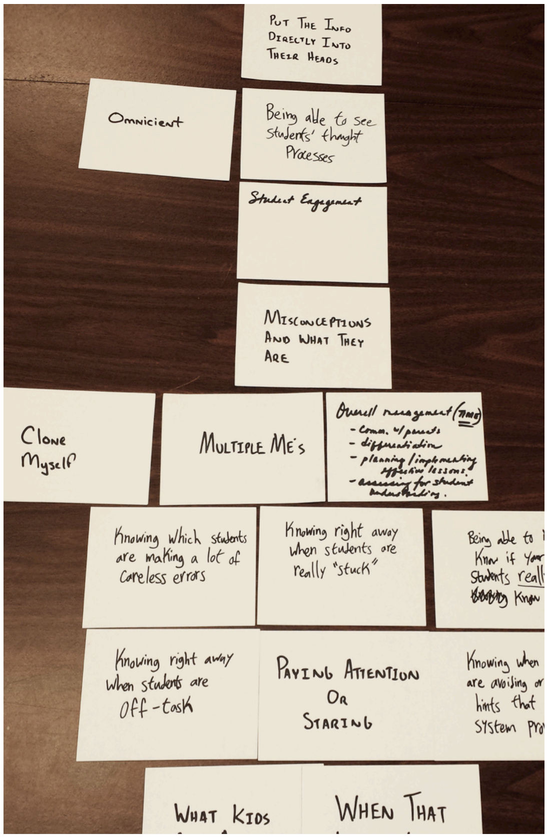

Figure 1. Hierarchy generated from a teacher's superpower card sort. The superpowers this teacher considered most desirable are at the top of this hierarchy. Multiple cards placed at the same level of the hierarchy represent ties (or superpowers the teacher considered synonymous).

Figure 2 summarizes teachers' aggregate preferences between pairs of superpowers, within a pairwise comparison matrix. In this figure, only superpowers that appeared in at least two teachers' hierarchies are shown. Overall, teachers tended to prefer seeing students' thought processes over most other superpowers. Interestingly, most teachers ranked seeing thought process over seeing misconceptions. Some teachers commented that if they

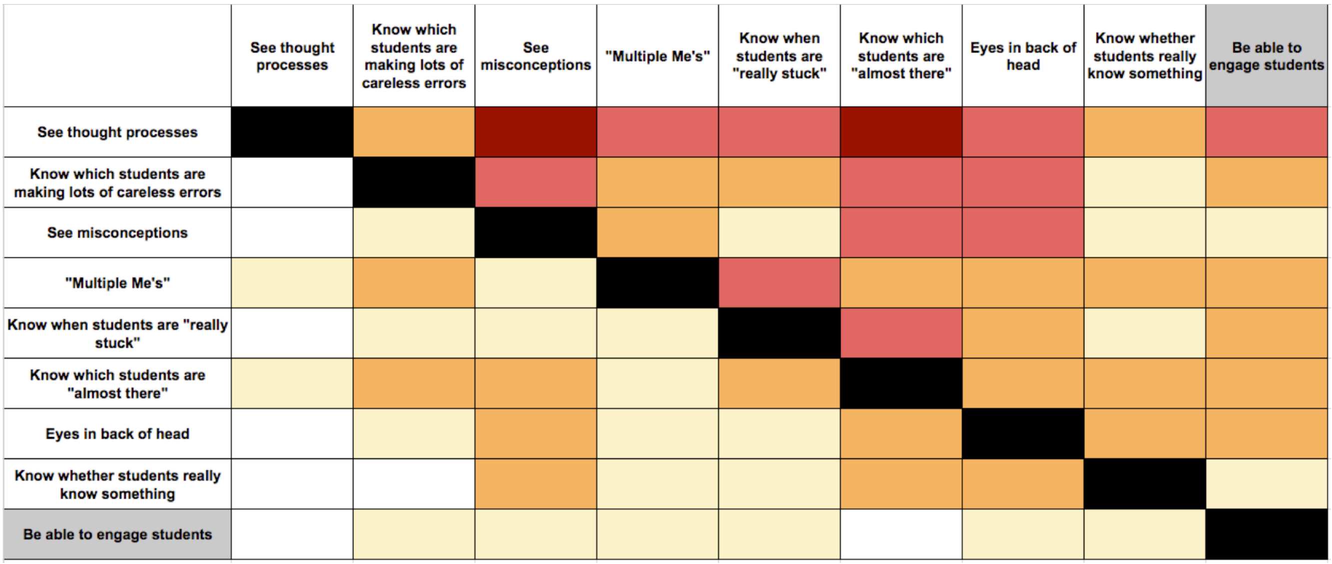

Figure 2. Pairwise comparison matrix summarizing teachers' preferences between superpowers. Each row shows a superpower that appeared in at least two teachers' hierarchies, and each column shows a different superpower against which it is being compared (cells on the diagonal represent self-comparisons, and are blacked-out). Cell shade indicates the number of teachers who ranked the row superpower higher than the column superpower, with darker shades indicating greater agreement (minimum observed value is 0, and maximum is 4). "Be able to engage students" is highlighted in grey to indicate that this superpower was not present in all five teachers' card stacks (by the time a teacher first generated this card, no synonyms were available among the cards generated by previous teachers).

could really see and understand students' step-by-step reasoning, that would likely reveal students' misconceptions *and more* – perhaps providing an explanation for this preference. It is also worth noting that *knowing whether students really know something* was not strongly favored by teachers overall, compared with other superpowers, despite estimates of student knowledge (e.g., in the form of probabilities that a student has mastered particular skills) being one of the most central analytics presented by common reporting systems for ITSs (e.g. [36]).

Across the hierarchies of superpowers that teachers generated, some interesting regularities emerged. All five of the interviewed teachers specified that they wanted the ability to:

See students' thought processes. Teachers wanted to be able to see students' thought processes – the chains of reasoning that led them from one statement to the next – without having to ask students to "show their work" (and without the need to subsequently interpret students' work and try to infer their underlying thought processes). Some teachers explicitly distinguished this from simply seeing estimates of students' mastery over certain skills (as was presented by adaptive learning technologies with which they were familiar), noting that they viewed seeing thought processes as more *actionable* than seeing skill estimates. If teachers could follow students' thought processes in real-time, this would provide opportunities to "re-route" students at the moment when they "took a wrong turn".

Know which students are *truly* stuck. Teachers noted that students often raise their hands during lab sessions when they don't actually need help (and may simply be trying to avoid doing their work). At the same time, teachers believed that many students who actually need help almost never raise their hands. Being able to see which students actually need the teacher's help, at any given moment, would enable the teacher to better prioritize help across students, and "fight the biggest fires first".

Know which students are "almost there", and just need a nudge to reach mastery. Teachers noted that one of the most fulfilling parts of their jobs is "seeing students to the finish line": working with students who are currently on the verge of understanding a new concept, and helping them reach that understanding more quickly. One teacher was initially conflicted over whether to include this superpower in his hierarchy, noting that these students would likely reach mastery even without the teacher's help, and that other students may need his help much more. But this teacher ultimately decided to keep this superpower, acknowledging that he wouldn't want to spend all of his time as a teacher focusing on the students who are struggling the most.

In addition, four out of the five of the teachers interviewed wanted to be able to:

Temporarily clone themselves (create "Multiple Me's"). Teachers wanted the ability to provide one-on-one support to many students simultaneously, rather than leaving instructional personalization entirely up to the software. All of the interviewed teachers highly value the instructional differentiation provided by such software, but also acknowledge that such differentiation makes it challenging for teachers to keep track of their students' current activities, let alone provide them with timely instruction.

Have "eyes in the back of my head". Teachers noted that certain students tend to take advantage of the challenges ITS lab sessions pose for classroom monitoring, by misbehaving specifically when the teacher's back is turned. They shared stories of middle school students switching to non-academic websites when they thought the teacher was not watching, but immediately switching back to the ITS interface when they knew they were in visual range. Thus, much of these teachers' energy during lab sessions is spent "patrolling" the room and trying to make sure everyone is on-task.

Detect students' misconceptions. Similar to teachers' desire to see students' thought processes, their wish to immediately diagnose students' misconceptions was rooted in the *actionability* of this information. While teachers viewed *seeing students'*

kids to tutor
er (peer tutors)

ing for the kids
successful with
dy. Because I can
to help the other
eer tutors"

d out that you have
who's just blowing
e water, they're
g well... how
ing him help other
who are having

ee who had
d on which level...
f you can pair
y strong with
y weak... they
better to each

ight give back the
and say now what
o wrong? Or look

Challenges of managing/running an effective classroom

"It takes so much time to run an effective classroom on a day-to-day now... every month, we're learning something new"

"So you could individualize more..you know everyone talks about individualizing more... and you have a class of ..you know...you could say 21 students. 42 minutes, each student can have 2 minutes of your time"

"and this class is tough because you, it's a large, it's a low class and it's large. So that's kind of why there's two of us in here with Mr. xxxxxxxx and I. Uhm, why we're both in here at the same time. Just because you have kids in all different levels, of all different

Importance of the teacher's presence, face-to-face interaction

"to motivate or help some students, i need to be there to talk to them"

"the only thing that's gonna help that below average student is when I get involved and it's one on one."

"When they see me come in and sit down, before they embarrass themselves by not being able to do what it is I want them to do, a lot of times they'll tell me 'Mrs. xxxxxxx, I'm really confused.'"

"often times it was just getting them to see where that was or what it was]..and then they were able to progress through it... if it was something like all the

Group work as a way to reduce teacher's load

"they can explain it to each other better than I can explain it to them... they'll ask their seat partner and then they'll turn and ask somebody else around them .. it saves me a whole lot of having to clear up little things where I can go and work with that one kid or 2 kids in the class who are completely lost the whole time"

"I'll be like, talk as a group and see if everybody can get on the same track, and then when you're on the same track I want you to raise your hand and call me over to make sure that track is the right on"

"Now a lot of times let's say, this independent work is challenging, well I might

Encouraging self-explanation

"I'm really big in having the students explain stuff because the more they can explain it then the better they are going to understand it... like sometimes I'll write down the wrong answer and I'll just keep going and you know I'm waiting for one of them to say 'wait minute. Why are we doing that?'"

"We've seen over the years that the writing aspect of it is the key.. if they are able to write down and explain how to do it, that's when we know that they have got it .... if they can't explain how to solve the various equation they just don't have it mastered"

"In some cases, what I try to do is if it's a mistake that I saw that the student had made. Sometimes I would

Figure 3. A partial view of our affinity diagram, showing teacher quotes within level-1 categories.

thought processes as a means to correct particular student errors and mold students' thinking in desired directions, they viewed *detecting students' misconceptions* as an opportunity to eliminate problematic beliefs that might hinder their future learning.

Know which students are making lots of careless errors. Finally, teachers wanted to be able to automatically detect whether students are actually putting in the effort required to learn. Based on this information, they could decide whether it would be most productive to spend their time on an *instructional* intervention, or whether it would be more appropriate to first determine how to motivate the student.

3.3 Eliciting and synthesizing teachers' design requirements for intelligent real-time supports

We conducted semi-structured interviews with 10 teachers from 5 schools (all teachers and schools identified in Tables 1 and 2), in order to more directly investigate teachers' needs for real-time supports. In these interviews, we asked teachers to reflect on their experiences using adaptive educational technologies such as ITSs in the classroom, and to consider how these and similar systems could be better designed for use in their classrooms. In particular, we encouraged teachers to imagine that the ITS could communicate with them in real-time, and that there were no limits on what the system could measure about student learning.

A PhD student (the first author of this paper) and a research assistant (an undergraduate design student from our institution) then worked through transcriptions of approximately 5 hours of video and audio recorded interviews, to analyze and synthesize findings from the interview data, using two standard techniques from Contextual Design: Interpretation Sessions and Affinity Diagramming. Interpretation Sessions aim to help design teams create a shared understanding of collected interview data by extracting quotes representing key issues and insights from each participant's interview. Affinity Diagramming is a widely used design method that aims to summarize patterns across the interviewed population by iteratively clustering and organizing interview quotes, based on content similarity, into a hierarchy of increasingly abstract themes [16].

Both the extraction of quotes during Interpretation Sessions and the hierarchical clustering process of Affinity Diagramming are inherently subjective processes. However, bias can be reduced during Interpretation Sessions by taking a conservative approach to quote extraction (i.e. erring on the side of including a large number of quotes, even when some of these may not seem particularly interesting or important). In addition, Affinity Diagramming involves a grounded, bottom-up approach to data analysis: higher-level categories gradually emerge during the clustering process via agreement across team members [18]. As such, the process is designed to minimize the extent to which the resulting summary is guided towards individuals' preconceptions, without entirely removing the influence of prior knowledge (e.g., knowledge of extracted quotes' context).

We conducted several Interpretation Sessions of approximately 3 hours of transcribed interviews. The resulting 301 quotes were then iteratively synthesized into higher-level categories, through Affinity Diagramming sessions. Due to the large number of extracted quotes, we opted to conduct these Affinity Diagramming Sessions digitally, using Trello[1]: a web-based, drag-and-drop tool that allows users to easily organize information.

Following the Affinity Diagramming method, we first organized quotes into level-1 categories – which were initially unnamed – based on perceived similarity in their content. We then synthesized the quotes within each level-1 category, labeling each category by a summary of its contents. Once we had labeled all level-1 categories, we proceeded to group these into unnamed level-2 categories, and then repeated the synthesis and labeling process described above for these higher-level categories. Finally, we repeated the grouping, synthesis, and labeling process for all level-2 categories, producing a set of fairly abstract level-3 "themes". The resulting Affinity Diagram (shown in Figure 3) had 40 level-1 categories (with between 1-2 to 20-27 quotes per category), 10 level-2 categories, and 4 level-3 categories.

The most common high-level themes (level-3) reflected the interviewed teachers' common desires to maintain control of their own classrooms, even when students are working with intelligent tutoring systems, and to be an effective force in the classroom, over and above what these technologies can offer. Both of these

---

[1] Trello is freely available at https://trello.com/.

desires were often accompanied by comments about teachers' perceptions that technologists intend to *replace* their role with educational technologies, instead of *supporting* their roles as teachers. In addition, the level-3 themes revealed a desire to receive analytics that can truly teach them something about either their students learning or their own teaching – a theme which was often accompanied by complaints about learning analytics dashboards and reporting systems that either provide them with information they are already likely to know, or provide them with data they find useless for informing their own pedagogy. Finally, one level-3 theme reflected teachers' concerns that real-time analytics in the classroom could, if not designed carefully, cause more harm than good.

Within these high-level themes, the issues teachers raised broke down into the following 10 mid-level categories:

Help me to intervene *where*, *when*, and *with what* I'm most needed. Teachers wanted support from the ITS in deciding how best to prioritize their time across multiple students who may compete for their attention at once, when to help (or not help) a given student, and how best to help individual students. Given teachers' limited time during lab sessions, recommendations about *how* to help students might come as quick advice about effective instructional strategies to use, or pointers to educational resources (e.g. targeted remedial materials, available online) that may be helpful to the student.

I'm just one person: help ease my load. Teachers emphasized the usefulness of group activities and peer tutoring in reducing their orchestration load in the classroom. They suggested that one way an ITS could help them during a lab session would be to recommend pairings or small groups of students that are likely to be able to help one another (perhaps adaptively matched by the ITS based on its knowledge of their current skills). This would remove some of the responsibility of helping students from the teacher's shoulders, and also provide opportunities for the teacher to work with a larger number of students (by meeting with groups rather than individuals).

How can I know whether what I'm doing is *actually working*? Teachers noted that they very rarely have opportunities for immediate feedback on their own teaching. They often worry, especially after seeing students' test scores, that much of what they have taught students over several weeks or months has had no effect. Observing that ITSs can already track aspects of student learning in real-time, they wanted ITSs to provide *them* with timely feedback on the effectiveness their own help-giving during lab sessions (e.g., one-on-one interactions with individual students, or targeted mini-lectures provided to the whole class). This would allow them to adjust their strategies on the fly.

Help me understand the "why", not just the "what". Given teachers' active and time-constrained role during ITS lab sessions, they wanted ITSs to provide them with summarized and directly actionable information whenever possible. Teachers noted that most reporting systems they had used in the past, for ITSs and other educational technologies, tended to provide data about students' raw actions in the software. But a real-time monitoring tool for use during ITS labs would need to provide diagnoses of problems the teacher could act upon. For example, rather than simply presenting teachers with the observation that a particular student is making frequent errors in the software, it would be highly valuable to also *help the teacher diagnose* whether this is due to carelessness or genuine struggles with the material (and if the latter, to also help the teacher diagnose what those struggles may be).

But how do I judge whether my students are *really* doing well? Teachers wanted support from the ITS in determining what constitutes "good" performance in an ITS (e.g., is a 70% probability of mastery below or above "average", for a particular skill). To provide benchmarks for evaluating their students' performance, teachers suggested that ITSs might share descriptive statistics about overall class performance between multiple classrooms and schools.

Help me manage student motivation and engagement in my classroom. Teachers wanted ITSs to provide them with real-time analytics about their students' motivation and affective states in the classroom, in addition to analytics about student learning and performance. Notifications about student frustration while working with the ITS, for example, could allow teachers to intervene before the situation worsened.

What can you tell me about my students that I don't already know? Teachers complained that reporting systems they had used in the past tended to provide them with a lot of unsurprising information about their students. They wanted ITSs to somehow take into account their rich prior knowledge about their students (e.g., "[this student] is going to make slower progress, but that's only because she's so deliberate"). In doing so, ITSs could then provide teachers *mainly with notifications that are likely to surprise them.*

Make sure that the technology does not contribute to an unhealthy classroom climate! Teachers worried that real-time learning analytics presented by an ITS could easily draw their attention away from their classroom, thus defeating the purpose. They emphasized that an effective classroom monitoring system would need to be designed to keep teachers eyes and ears on the classroom to the greatest extent possible.

Let me customize the technology to meet my needs. Teachers noted that ITSs often seem rigid and inflexible. In cases where the instructional design of an ITS does not align with their own pedagogy, teachers want to be able to quickly and easily adapt the ITS to fit their needs.

Allow me to take control of the intelligent tutoring system. Teachers also wanted to be able to go a step beyond customization by actually taking control of the ITS on-demand. For some teachers, this simply meant being able to "pause" all of their students' screens while giving a lecture in the midst of a lab session, in order to ensure the software would not compete with the teacher for their attention. For other teachers, this meant being able to load a "quiz problem" on all students' screens (in order to quickly assess the whole class on a targeted set of skills).

4. SPEED DATING POSSIBLE FUTURES

In order to further probe and validate teachers' needs for real-time supports in ITS classrooms, we presented teachers with futuristic classroom scenarios – inspired by teachers' own ideas, as elicited through the superpowers card sorting exercise and our semi-structured interviews. We adopted "Speed Dating", a design method for rapidly exploring new technology concepts, in which participants are presented with a number of imagined futures in quick succession (represented through technology usage scenarios, in concept sketches and storyboards), while researchers observe and explore their gut reactions to these futures [5, 6]. Speed Dating allows designers and researchers to probe the

Figure 4. Panels from two storyboards used in Speed Dating sessions. Left: a teacher wears augmented reality glasses, which help her identify which of her students most need her attention in class. When the intelligent tutoring system detects that a student needs help from a human teacher, it "raises the students hand for them" – visible only within the teacher's augmented reality display. The ITS detects that a student on the far left of the image is browsing a non-academic webpage, and alerts the teacher through her augmented reality glasses by displaying a "Zzz…" over that student's head. Right: an extreme example of a concept that we did not expect teachers would find acceptable. In this storyboard, the teacher can control a drone inside the classroom, which collects additional data on students, and reminds students that the teacher is watching them. Each student is equipped with various biosensors, and aspects of their physiological and affective states are displayed to the teacher in real-time.

boundaries of what participants find acceptable (which otherwise often go undiscovered until after a prototype has been built and tested, or even after a technology has been deployed), by presenting them with imagined scenarios that are expected to cross these boundaries. This method can also lead to the discovery of unexpected design opportunities, when anticipated boundaries are found not to exist, or when unexpected needs are revealed. Importantly, Speed Dating can often reveal surprising results about a user population's needs and desires, above and beyond what methods for design requirements elicitation, like those described above, can reveal.

We met again with five teachers from our previous interviews (teachers 4, 7, 8, 9, and 10, from schools C and D), and presented them with futuristic classroom scenarios – inspired by teachers' own ideas, as elicited through the superpowers card sorting exercises and our semi-structured interviews. Teachers were presented with eleven storyboards; each presenting a futuristic scenario based upon a combination of teachers' most commonly requested superpowers and the 10 mid-level categories derived from our affinity diagram. We next summarize some key findings from these Speed Dating sessions below.

Contrary to teachers' expressed desire for real-time support in prioritizing their time across multiple students during a lab session, teachers consistently rejected the concept of "time management" systems that remind teachers not to spend "too much" time with students who are doing well without the teacher's help (instead directing them towards other students who may benefit more from assistance). One teacher reacted strongly, stating, "I don't need that… to remind me it's time to move on. I know that. As an educator, you know when you've got other kids to deal with". Although recent work suggests that, contrary to this teacher's assertion, educators' intuitions about which students need the most help (and when to help) may be limited [41], this teacher's comment reflects a key tension in the design of teacher support tools. Teachers' comments in response to this storyboard suggested that such alert systems are undesirable both because they threaten teachers' autonomy in the classroom, and also because they remove teachers' ability to choose between two of their primary desires during lab sessions (as indicated by their requested superpowers): helping the students who are struggling the most versus helping students who are "almost there".

By contrast, teachers were highly receptive to technology designs that presented them with information that could help them prioritize their time among students, without attempting to *directly recommend* certain actions. For example the panel on the left in Figure 4 is from a storyboard showing an augmented reality glasses based monitoring tool, through which an ITS can inform the teacher that a given student may need their assistance (even if the student is not necessarily aware that she/he needs help). One teacher noted that such a tool would be particularly helpful because "there are always students who are shy and just don't raise their hands… [and some] raise their hands when they really don't need help". A key reason teachers liked the concept of an augmented reality glasses based dashboard was that, by displaying analytics directly overtop their field of view, this technology would not draw their attention away from the classroom. In fact, teachers' reactions to this and other storyboards (including a smart watch based real-time dashboard) suggest that they may be quite open to, and in fact *strongly prefer* wearable learning analytics displays to handheld displays such as mobile phones and tablets. Such wearable displays hold the potential to minimize teachers' cognitive load as they move throughout the classroom, while also maintaining students' privacy. In particular, teachers saw a head-mounted, augmented reality display as an opportunity to have their own "private" smart classroom: one that only they could see.

5. DISCUSSION AND FUTURE WORK

In this paper we investigate how intelligent tutoring systems could be better designed to meet teachers' needs in K-12 blended classrooms, and we identify several design opportunities for real-time teacher support tools in ITS classrooms. Through semi-structured interviews with middle school teachers, we identified opportunities for ITSs to better support teachers, for example, in fairly and accurately assessing their students' performance within the software. In addition, these interviews suggested potentials for *predictive* analytics to aid teachers in making challenging decisions, such as whether to override ITSs' built-in mastery learning algorithms in order to keep slower-moving (perhaps struggling) students in pace with the rest of the class.

Through card sorting exercises and design interviews, we identified ITS design features and requirements for real-time analytics that may help address some of teachers' greatest challenges in ITS classrooms. Importantly, these findings suggest that the analytics most commonly generated by existing reporting systems for ITSs do not align with the analytics that teachers expect to be most useful and *actionable*.

By testing a number of alternative futuristic scenarios with teachers, using the Speed Dating design method, we discovered that K-12 teachers were highly receptive to the concept of intelligent classroom monitoring tools that support them in deciding how best to allocate their time and attention across students during a lab session. However, they strongly disliked the idea of such a system providing explicit, unsolicited recommendations for action. We do not interpret these findings to mean that teacher support tools should avoid making clear recommendations for action. Indeed, given previous findings that teachers sometimes make decisions that are suboptimal or even harmful to students' learning with ITSs (e.g., [31]), we suspect that such directness could be important in guiding teachers towards more effective interventions – perhaps especially in real-time usage scenarios, where teachers have scarce time to pore over data visualizations and draw their own conclusions. Rather, we believe these findings highlight a delicate tension between technology designers' desire to "nudge" teachers towards instructionally effective patterns of behavior, on the one hand, and the need to privilege teachers' autonomy and rich prior knowledge, on the other (paralleling recent findings in other domains where intelligent systems are developed to support human experts' decision-making, such as healthcare [42]). It may be, for example, that teachers would be more receptive to more explicit and direct action recommendations if these were presented only upon a teacher's request, rather than in the form of automated alerts. This question, and the broader question of how teacher support tools can achieve an effective balance between simply augmenting teachers' *awareness* in the classroom [24, 40] and more directly supporting their *decision-making* [22, 39] remain interesting open questions for future design and experimental work.

Our findings provide novel insights into teachers' needs for real-time support tools in classrooms using intelligent tutoring systems. In addition, to the best of our knowledge, this is the first study that presents a broad exploration of K-12 teachers' needs for real-time learning analytics in ITS classrooms. These findings may be useful for designers of adaptive educational technologies, as well as designers of real-time monitoring tools such as dashboards. We expect that many of our findings regarding teachers' needs for real-time analytics may generalize to a broader class of educational software than ITSs.

The next phase of our project involves the use of these results to inform the design of a real-time dashboard for K-12 teachers using ITSs, as a first step towards designing intelligent tutoring systems that can effectively collaborate with human teachers. While the findings presented in this paper provide much direction for design and highlight key teacher needs that a particular design may meet or fail to meet, it is nonetheless a challenging design problem to build a real-time dashboard that can both serve teachers' needs and ultimately enhance student learning in ITS classrooms. Given our findings from Speed Dating, we plan to explore the viability and affordances of wearable technologies (such as smart watches or augmented reality glasses) as real-time monitoring tools for K-12 teachers. Continuing our user-centered design process, we will run 'simulated lab sessions' in which teachers experience using multiple alternative prototype designs in a simulated classroom setting. In particular, we plan to use these sessions to test and validate real-time analytics that are designed to meet some of teachers' requests for superpowers.

6. ACKNOWLEDGMENTS

This work was supported by NSF Award #1530726, and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. In addition, we thank Jasper Tom, Franceska Xhakaj, Mary Beth Kery, and all participating teachers and students.

7. REFERENCES

[1] Matuk, C., Gerard, L., Lim-Breitbart, J., & Linn, M. (2016). Gathering requirements for teacher tools: Strategies for empowering teachers through co-design. *Journal of Science Teacher Education*, *27*(1), 79-110.

[2] Martinez-Maldonado, R., Dimitriadis, Y., Kay, J., Yacef, K., & Edbauer, M. T. (2013). MTClassroom and MTDashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. In *Proc. CSCL2013*, 119-128.

[3] Xhakaj, F., Aleven, V., & McLaren, B. M. (2016). How Teachers Use Data to Help Students Learn: Contextual Inquiry for the Design of a Dashboard. In *European Conference on Technology Enhanced Learning*, 340-354. Springer International Publishing.

[4] Nye, B. D. (2014). Barriers to ITS adoption: A systematic mapping study. In *ITS 2014,* 583-590. Springer International Publishing.

[5] Davidoff, S., Lee, M. K., Dey, A. K., & Zimmerman, J. (2007). Rapidly exploring application design through Speed Dating. In *International Conference on Ubiquitous Computing*, 429-446. Springer Berlin Heidelberg.

[6] Hanington, B., & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.

[7] Yacef, K. (2002). Intelligent teaching assistant systems. In *ICCE 2002*, 136-140. IEEE.

[8] Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the Assistment system. *Technology Instruction Cognition and Learning*, 3, 1-8.

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

[10] Kinshuk, Tretiakov, A., Hong, H., & Patel, A. (2001). Human teacher in intelligent tutoring system: a forgotten

entity. In *Proc. ICALT 2001*, 227-230. IEEE Computer Society.

[11] Vivet, M. (1992). Uses of ITS: Which role for the teacher?. In *New Directions for Intelligent Tutoring Systems*, 171-180. Springer Berlin Heidelberg.

[12] Chen, Z. (1991). From student model to teacher model: Enriching our view of the impact of computers on society. *SIGCAS Computers and Society*. 21(2-4), 46-48. ACM.

[13] Tissenbaum, M. & Matuk, C. (2016). Real-time visualization of student activities to support classroom orchestration. In *ICLS 2016,* 1120-1127.

[14] Cairns, P., & Cox, A. L. (Eds.). (2008). *Research methods for human-computer interaction* (Vol. 12). New York (NY): Cambridge University Press.

[15] Pinkwart, N. (2016). Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *IJAIED*, 26(2), 771-783.

[16] Beyer, H., & Holtzblatt, K. (1997). Contextual design: defining customer-centered systems. Elsevier.

[17] Evenson, S. (2006). Directed storytelling: Interpreting experience for design. *Design Studies: Theory and research in graphic design*, 231-240.

[18] Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.

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

[20] Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., Ritter, S., Knight, A., Mercado, E., Turner, T., Upalekar, R., Walonoski, J., Macasek, M., Aniszczyk, C., Choksey, S., Livak, T., & Rasmussen, K. (2005). The Assistment project: Blending assessment and assisting. In *AIED* 2005, 555-562.

[21] McLaren, B.M., Scheuer, O., & Mikšátko, J. (2010). Supporting collaborative learning and e-Discussions using artificial intelligence techniques. *IJAIED* 20(1), 1-46.

[22] Vatrapu, R., Teplovs, C., Fujita, N., & Bull, S. (2011). Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In *LAK 2011*, 93-98. ACM.

[23] Prieto, L. P., Sharma, K., & Dillenbourg, P. (2015). Studying teacher orchestration load in technology-enhanced classrooms. In *Design for Teaching and Learning in a Networked World*, 268-281. Springer International Publishing.

[24] Rodriguez Triana, M. J., Prieto, L. P., Vozniuk, A., Shirvani Boroujeni, M., Schwendimann, B. A., Holzer, A. C., & Gillet, D. (in press). Monitoring, Awareness and Reflection in Blended Technology Enhanced Learning: a Systematic Review. *IJTEL*.

[25] Lesta, L., & Yacef, K. (2002). An intelligent teaching assistant system for Logic. In *ITS* 2002, 421-431. Springer Berlin Heidelberg.

[26] Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In *New Science of Learning*, 525-552. Springer New York.

[27] Prieto, L. P., Sharma, K., Dillenbourg, P., & Jesús, M. (2016). Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In *LAK 2016*, 148-157. ACM.

[28] Segedy, J., Sulcer, B., & Biswas, G. (2010). Are ILEs ready for the classroom? Bringing teachers into the feedback loop. In *ITS 2010*, 405-407. Springer Berlin Heidelberg.

[29] Miller, W. L., Baker, R. S., Labrum, M. J., Petsche, K., Liu, Y. H., & Wagner, A. Z. (2015). Automated detection of proactive remediation by teachers in Reasoning Mind classrooms. In *LAK 2015*, 290-294. ACM.

[30] Matuk, C., Gerard, L., Lim-Breitbart, J., & Linn, M. C. (2016). Teachers' reflections on the uses of real-time data in their instruction. Poster session presented at *AERA 2016*, Washington, DC, USA.

[31] Ritter, S., Yudelson, M., Fancsali, S. E., & Berman, S. R. (2016, April). How Mastery Learning Works at Scale. In *L@S 2016*, 71-79. ACM.

[32] Käser, T., Klingler, S., & Gross, M. (2016). When to stop?: towards universal instructional policies. In *LAK 2016*, 289-298. ACM.

[33] Liu, R., Patel, R., & Koedinger, K. R. (2016). Modeling common misconceptions in learning process data. In *LAK 2016*, 369-377. ACM.

[34] Baker, R.S.J.d, Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., & Rossi, L. (2012). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In *EDM* 2012, 126-133.

[35] Hardy, M. E. (2004). Use and evaluation of the ALEKS interactive tutoring system. *Journal of Computing Sciences in Colleges*, 19(4), 342-347.

[36] Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249-255.

[37] Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R., Aleven, V., Cocea, M., Hershkovitz, A., de Carvalho, A.M.J.B., Mitrovic, A., Mathews, M. (2013). Modeling and Studying Gaming the System with Educational Data Mining. In Azevedo, R., & Aleven, V. (Eds.) *International Handbook of Metacognition and Learning Technologies*, 97-116. New York, NY: Springer.

[38] Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253-278.

[39] Borko, H., Roberts, S. A., & Shavelson, R. (2008). Teachers' decision making: From Alan J. Bishop to today. *In Critical Issues in Mathematics Education*, 37-67. Springer US.

[40] Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Routledge.

[41] Holstein, K., McLaren, B.M., & Aleven, V. (in press). SPACLE: Investigating learning across virtual and physical spaces using spatial replays. In *LAK 2017*. ACM.

[42] Yang, Q., Zimmerman, J., Steinfeld, A., Carey, L., & Antaki, J. F. (2016). Investigating the Heart Pump Implant Decision Process: Opportunities for Decision Support Tools to Help. In *CHI 2016*, 4477-4488. ACM.