Effects of a Teacher Dashboard for an Intelligent Tutoring System on Teacher Knowledge, Lesson Planning, Lessons and Student Learning

Françeska Xhakaj, Vincent Aleven, Bruce M. McLaren

Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA
francesx@cs.cmu.edu, aleven@cs.cmu.edu, bmclaren@cs.cmu.edu

Abstract. Intelligent Tutoring Systems (ITSs) help students learn but often are not designed to support teachers and their practices. A dashboard with analytics about students' learning processes might help in this regard. However, little research has investigated how dashboards influence teacher practices in the classroom and whether they can help improve student learning. In this paper, we explore how Luna, a dashboard prototype designed for an ITS and used with real data, affects teachers and students. Results from a quasi-experimental classroom study with 5 middle school teachers and 17 classes show that Luna influences what teachers know about their students' learning in the ITS and that the teachers' updated knowledge affects the lesson plan they prepare, which in turn guides what they cover in a class session. Results did not confirm that Luna increased student learning. In summary, even though teachers generally know their classes well, a dashboard with analytics from an ITS can still enhance their knowledge about their students and support their classroom practices. The teachers tended to focus primarily on dashboard information about the challenges their students were experiencing. To the best of our knowledge, this is the first study that demonstrates that a dashboard for an ITS can affect teacher knowledge, decision-making and actions in the classroom.

Keywords: intelligent tutoring systems, dashboard, data-driven instruction, teachers' use of data, learning analytics

1 Introduction

Intelligent Tutoring Systems (ITSs) are a type of advanced learning technology that provides detailed guidance to students during complex problem-solving practice, while also being adaptive to student differences [3, 21, 24]. ITSs have been shown to enhance student learning [8, 11, 19]. However, ITSs are rarely designed to support teachers, who might greatly influence student learning with an ITS. The addition of a teacher dashboard might help them do so. For instance, when many students in a class are learning a particular skill as they are working with the ITS, a dashboard could let the teacher know about this situation, and the teacher could include, in their lesson plan and actual lesson, specific steps to address the challenge. More generally, a

dashboard could help make "the invisible visible" for teachers by displaying aggregated, up-to-date information about their students. Based on this information, teachers could provide help to their students beyond what the ITS can provide.

By now, researchers have developed many dashboards with analytics from educational technologies. Much research focuses on evaluating whether such dashboards are useful to teachers and what visualizations or information is most used by them. Some studies found that a dashboard can help teachers determine in real-time when to intervene and help students work more collaboratively in a multi-tabletop learning environment [13], or can help them single out problems concerning participation in digital discussion environments and intervene as needed [20]. Other studies have shown that a dashboard's information can help teachers manage web-based distance courses [15], support teachers in moderating discussions in digital learning environments [16] or support their awareness of the classroom state, student progress, and students in need of immediate help in an exploratory learning environment [14].

In the current work, we focus on creating a teacher dashboard for an ITS, in contrast to much other research on dashboards. Given the somewhat unique characteristics of ITSs, it seems reasonable to assume that a dashboard for ITSs would be different compared to dashboards for other learning technologies. ITSs generate and collect data related to self-paced learning with step-level support for problem solving, adaptive mastery learning based on a detailed skill model, characteristics not widely shared with other educational technologies. In addition, ITSs typically generate and maintain a student model, which might create some interesting opportunities for dashboards. Exceptions are work by Lovett et al. (2008) who report on instructors using reports from an ITS in an online course [10], by Arroyo et al. (2014) who describe teacher reports generated by an ITS [4], and by Kelly et al. (2013) who study how a teacher used a report from a web-based homework system to decide what parts of the homework to review in class [7].

Further, while much work has focused on real-time dashboards (dashboards that teachers use while students are working with a learning software in class), few have looked at other scenarios in which a dashboard might be helpful. In the current work, we look at a scenario in which a teacher uses a dashboard when preparing for a class session; a dashboard might help in focusing the class discussion on the topics most in need of discussion (e.g., problems or specific error types that are currently challenging for the students). One study that comes close to this scenario is Kelly et al. (2013) who found positive effects of in-class review of reports from a web-based homework system [7]. In another study, Mavrikis et al. (2015) report that information from a dashboard about difficulties students are facing in an exploratory learning environment may help teachers decide what to focus on in the following lesson [14].

Finally, although many evaluation studies involving dashboards have been conducted, few studies have looked at the influence a dashboard might have on student learning, in spite of a growing realization in the field that effects on student learning should be studied [18, 22]. In the current paper, we present results from a quasi-experimental classroom study investigating effects of a dashboard prototype, Luna, with analytics from an ITS, used for lesson planning. Our study looks at effects on teacher knowledge, teacher decision-making, and student learning. It looks at realistic

decision making, namely, planning and executing a classroom lesson following sessions during which the students used the ITS.

2   A Causal Chain that Captures Dashboard Influences

We defined a hypothesized causal chain that represents how information in a dashboard may affect teachers and, through them, student learning (Figure 1). It focuses on scenarios in which a teacher uses a dashboard to prepare for a class session, in blended courses that use some form of educational technology. The dashboard, it is assumed, displays up-to-date information about students' performance, progress, and learning, with some technology. The causal chain may apply to any dashboard, learning analytic tool, teacher awareness tool, or report on student learning in blended courses, where teachers use it to create a lesson plan and prepare for a class session.

From their experience with a particular class, teachers have knowledge about what their students generally can and cannot do well, at any given point in time (link 1, Figure 1). As they work with a dashboard, they may learn new information about the performance and knowledge of their students (link 2 in Figure 1). When teachers plan for a class session, their updated knowledge may affect the lesson plan (link 3 in Figure 1), which then guides what they cover in class (link 4 in Figure 1). Ultimately, what teachers do in the class session is what students get exposed to and what affects their learning (link 5 in Figure 1). Thus, the dashboard information needs to "travel" through many links; it must be embraced by teachers, incorporated in the lesson plan and used in the class session, for it to reach students and impact their learning. In our analysis, we investigate the dashboard's influence along each of the links in the chain.
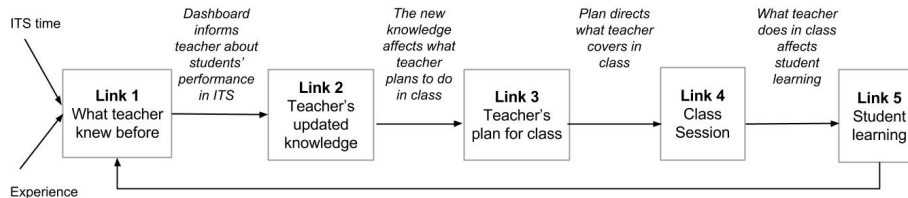
Fig. 1. A causal chain that represents a dashboard's effects on teacher practices.

This causal chain differs from the LATUX [12] framework, which describes ways to design, develop, evaluate and deploy learning analytics tools for teachers. By contrast, the causal chain captures potential effects of a dashboard from proximal influences on teacher classroom practices and to distal influences on student learning.

3   Methodology

In this work, we focus on the following research questions: (RQ1) How does a dashboard with analytics from an ITS affect teachers' lesson planning and (subsequent) classroom sessions? and (RQ2) Does the teacher's use of the dashboard help students learn better? This early, preliminary evaluation is a formative evaluation. A key goal is to gather information that helps us in the redesign of the dashboard.
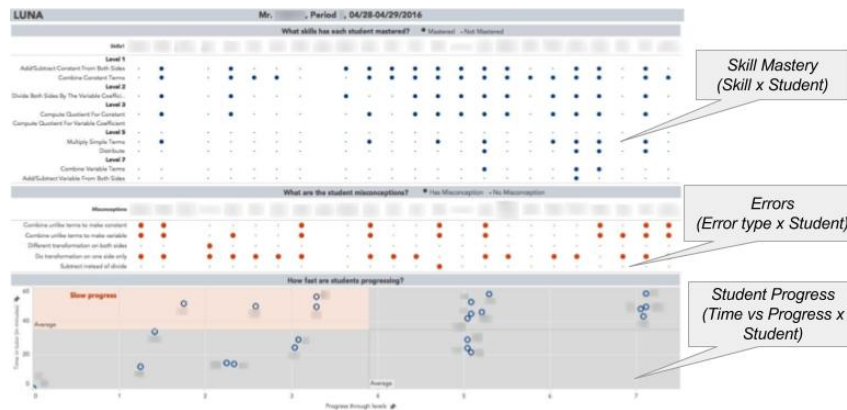
**Fig. 2.** Individual level dashboard prototype (Luna). Student names are obfuscated.

3.1 The Dashboard: Luna

Our study focused on Luna, a high-fidelity dashboard prototype (Figure 2). We created Luna employing a user-centered design approach [2, 6, 25]. We involved teachers in the design process through a variety of design methods including Contextual Inquiry, Speed Dating, Storyboarding and Prototyping [5]. Luna is powered with data from Lynnette, an ITS for middle school mathematics (grades 6-8) created with CTAT [1] and with an evidence-based record of helping students learn to solve linear equations [9, 23]. We used Tableau, a data visualization tool (http://www.tableau.com/), to create Luna's interface. In our study, we populated Luna with student data logged by Lynnette from the participating teachers' own classes. Luna displays data about students' learning, both at the class and individual level. At the class level, Luna shows (1) the number of students who have mastered each skill in Lynnette (as a horizontal bar chart), (2) the number of students who made certain errors (as a horizontal bar chart), and (3) a comparison of the level of mastery versus the amount of practice per skill averaged across students (as a scatter plot). At the individual level (Figure 2), Luna shows per student (1) if they mastered each skill in Lynnette and the percent mastery, (2) if they had errors and the number of times they made each error, and (3) time versus progress in the ITS (as a scatter plot). Luna is interactive, for example hovering over a skill or error shows a definition and an example exercise of the skill being applied or the error manifesting. The Cognitive Mastery algorithm in Lynnette generates skill mastery information (essentially, the tutor's student model), while an extended cognitive model generates error types.

3.2 Experimental Design

Five teachers from two suburban schools took part in the study (17 classes, 300 students total). All classes were 7th grade (medium achieving or honors) except for a 6th grade honors class and an 8th grade low-achieving class. Two out of the five teachers

had participated in previous iterations of Luna's design. The experiment had two conditions, an experimental condition, in which teachers used Luna while preparing a lesson plan, and a control condition, in which there was no dashboard. Classes were assigned to conditions such that each teacher had classes in both conditions. Conditions were balanced per teacher and school in terms of the level of achievement (high or low achieving class) and the order in which they happened during the school day. There were 9 classes in the control condition and 8 in the experimental condition.
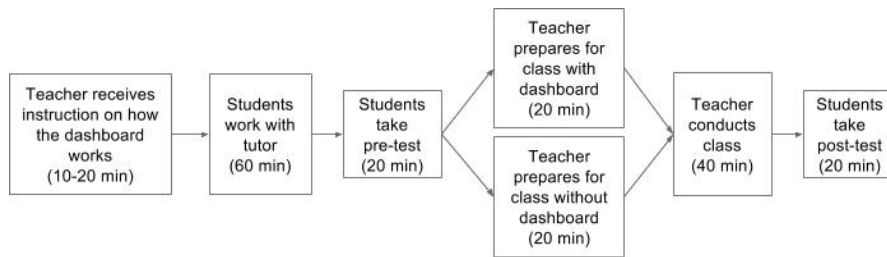
Fig. 3. Experimental set up for an individual teacher and an individual class.

We first provided teachers with 10-20 minutes of instruction on the analytics and visualizations that Luna displays (see Figure 3). For this instruction session, Luna displayed student data collected in previous studies. Students then worked for 60 minutes with Lynnette, completing problem sets dealing with basic equation solving. Next, they took a 20-minute pre-test. In both conditions, teachers were asked to prepare for 20 minutes for a class session and think out loud during the process; during these sessions, the researcher occasionally asked teachers to explain what they were doing. The sessions were video-recorded. For the experimental condition classes, teachers were asked to prepare for the class session using Luna, which provides information about their students' performance during the session with Lynnette. For the control condition classes, teachers were asked to prepare without a dashboard, based on their experience, their knowledge of their students, and on what they noticed when students were working with Lynnette in the lab. (The only difference between the two conditions therefore was whether or not the dashboard was available during the preparatory sessions.) Teachers then conducted the class sessions they prepared for. (The students did not use Lynnette during these sessions.) During these sessions, each 40 minutes, 2-4 coders (undergraduate students and staff from our institution) took observational notes using a tool with predefined categories of observations that also allowed for free-form note taking. After the class session, students took a 20-minute post-test. Both pre- and post-tests contained 9 exercises based on 9 problem sets in Lynnette, covered the same equation types, with different numbers, and were assigned in counterbalanced manner. The pre- and post-tests allow us to assess student learning gains due to the class session teachers conducted based on their preparation with or without the dashboard. (Learning gains due to the ITS would have happened prior to the pre-test.)

3.3 RQ1: How Does the Dashboard Affect Teachers?

We study how the dashboard affects teachers in each of the links of the causal chain.

Table 1. Table of a teacher's updated knowledge at the class and individual level.

Row		Code	Statement
1	Teacher's knowledge or expectations for class	--	KC1- Expect students are good at because they have done this already: Add/Subtract Constant From Both Sides
		--	KC2- Expect students are good at because they have done this already: Combine Like [Constant] Terms
		--	KC3- Expect students are good at because they have done this already: Divide Both Sides By Variable Coefficient
		--	KC4- The Distributive Property, I thought they would struggle with
2	Learned from dashboard for class	(✔)G	LC1- Add/Subtract Constant From Both Sides
		(✔)G	LC2- Combine Like [Constant] Terms
		(✔)G	LC3- Divide Both Sides By Variable Coefficient
		(+)G	LC4- Compute Quotient For Constant (8 did not get to Level 3, 16 who got there mastered it), ok that's good
		(+)N	LC5- 8 students did not get to Level 3
		(+)N	LC6- Combine Like [Variable] Terms (who got there mastered it, it's just that not everybody got there)
		(+)N	LC7- Add/Subtract Variables On Both Sides, the same kids who got to that [Combine Variable Terms] got this
		(✔)B	LC8- Distribute Property, ok that is where they are starting to fall of
		(+)N	LC9- A couple of kids did not grab this [gesturing Level 1 Add/Subtract Constant From Both Sides and Combine Constant Terms]
		(+)N	LC10- A couple of kids did not grab this [gesturing Level 2 Divide By Variable Coefficient]
3	Teacher's knowledge or expectations for individual students	--	KS1- Student 1 would be in one of the higher levels if she was here the first day
		--	KS2- Student 2 wasn't here at all
		--	KS3- Student 3 was here only the second day
		--	KS4- Student 4 and Student 5 would goof around if they work together
		--	KS5- Student 6 would be ok working with Student 7
		--	KS6- Student 4 is pretty strong
4	Learned from dashboard for individual students	(+)N	LS1- I have a high [level 7], medium [level 5], and low group [level 3]
		(!)N	LS2- Student 8 is kind of surprising
		(+)B	LS3- Student 1 is behind
		(+)B	LS4- Student 2 is at (0:0) (wasn't here?)
		(+)B	LS5- Student 3 is at (0:0) (thought was here the second day?)

Teacher's updated knowledge. Targeting the first link in the causal chain, we analyzed the video-recordings of the teachers' preparation sessions to assess how Luna affected their knowledge. From these video-recordings, the first author distilled and paraphrased the main ideas teachers expressed (which we will call statements) as they were thinking out loud during the preparation sessions. A second coder verified the segmentation of the recording into statements by time-tagging each of them. As shown in Table 1, we distinguished four categories of teacher knowledge, characterized by whether they knew it *before* inspecting Luna or became aware of it *while* inspecting it, and whether the focused-on information pertains to the class overall or to individual students. We created such tables with teachers' statements for each of the 8 experimental condition classes.

The statements that represent what teachers learned from the dashboard (rows 2 and 4 in Table 1) were coded based on two coding schemas. The first set of codes aims to classify how Luna's information relates to the teacher's prior knowledge, using the following codes: (1) "✔" means that Luna's information confirms what teachers knew about their students (e.g., *"Yeah, [student name] is not surprising…"*), (2) "**!**" means that teachers were surprised by Luna's information, or it was incon-

sistent with what teachers knew (e.g., *"The only thing that stands out for me is this [pointing at combine like terms make constant and make variable]…)*, and (3) "**+**" means that teachers learned from Luna, but it did not confirm or reject what they already knew, (e.g., *"… looking at it, [the]distributive property they have all pretty much mastered…"*). The second set of codes aims to classify whether the teacher's comment was about students doing well or not in Lynnette, based on data from Luna. It has the following codes: (1) "**G**" means that the teacher's comment is about information from Luna that showed students did well in Lynnette (e.g., *"I am actually kind of surprised that [student name] made it that far, that's good!"*), (2) "**B**" means that the teacher's comment is about students not doing well (e.g., *"… I see that that's what students have most trouble in, combine unlike terms to make a variable…"*), and (3) "**N**" means that the teacher's comment is ambiguous (e.g., if the teacher says, *"Only one hasn't mastered the distributive property,"* it is not clear whether he/she views that as positive or negative). The codes were assigned based only on what teachers explicitly said in the video-recordings of the preparation sessions. The first author and a trained coder first coded all statements independently. They then met and resolved all disagreements in coding through discussion and mutual consensus. The results reported here are based on this consensus coding.

Table 2. Lesson plan, with information attributable to Luna coded in the first column.

Code		Concepts teacher will cover/review in class (WHAT?)		Exercises teacher will do in class (HOW?)		
--		(Revise concepts through equation solving + students working in groups)				
LC9	1	Add/Subtract Constant From Both Sides		$x+8=-15$		
LC10	2	Divide Both Sides By The Variable Coefficient		$3x = 24$		
--	3	Distributive Property/Combining Like Terms				
LC8		a	(Stus started to fall of at the Distributive Property)	$5(x+4)=40$		
				$3=7(4-2u)-6u$		
LC6		b	(This has that combine in it)	$3(1+4n)-2(5n-3)=25$		
(From other class)	4	Variables On Both Sides				
		a	Variables On Both Sides	$5x+6=2x+15$		
		b	With Negative Numbers	$-7x-2=24-9x$		
		c	Distributive Property + Variables On Both Sides	$4(5n-7)=10n+2$		
		d	Distributive Property + Variables On Both Sides	$2(6d+3)=18-3(16-3d)$		
LS1, LS2, LS3, LS4, LS5	5	Students work in groups of 3 with worksheet with exercises on Distributive Property + Variables On Both Sides (same worksheet as previous class)	Level 7 1 Student 6 2 Student 9 3 Student 10 4 Student 11 5 Student 12 6 Student 13 7 Student 14 8 Student 4	Level 5 6 Student 15 3 Student 16 2 Student 17 1 Student 18 5 ~~Student 19~~ 7 Student 20 4 Student 21	Level 3 3. ~~Student 8~~ Student 24 1 Student 7 2 Student 22 5 Student 1 6 Student 23 7 Student 5 8 ~~x Student 3~~ x Student 2 (with Student 4 who is pretty strong and they were not here)	
--	6	Give worksheet from previous class				

Lesson Plan. Moving to the next link in the causal chain (link 3 in Figure 1), we analyzed how the knowledge gained from the dashboard may have influenced teachers' lesson plans. We focused on the lesson plans for the 8 classes in the experimental condition, which teachers created with help from Luna. To represent the lesson plans, we created tables (Table 2) based on the distilled and paraphrased main ideas teachers

mentioned or wrote down during the preparation sessions. These tables show the topics along with the exercises (if any) that teachers planned to cover during the class session, as well as their plans about individual students, when applicable. To study how the information learned from Luna affected the teacher's lesson plan, each of the items in the lesson plan (rows in Table 2) was matched with what teachers learned from Luna (rows 2, 4 in Table 1). For example, if the teacher stated, "… *that is where they are starting to fall off, at the distributive property*" (LC8 in Table 1) and then said *"… we are back into distributive property… so I can steal some examples from my other… [the plan for my other class] (writes down some exercises with the distributive property used in the previous class they prepared for),*" we would put the code LC8 under the respective row in the lesson plan table. This coding procedure was applied only to statements for which teachers explicitly stated that the reason they were going to cover it in class because was information from Luna.

Table 3. Part of a lesson plan compared with what happened during the class session.

Code		Concepts teacher will cover/review in class (WHAT?)	Exercises teacher will do in class (HOW?)	Covered?	Concept/Misconceptions	Who?
--		(Revise concepts through equation solving + students working in groups)		Yes	-Focus on what they will do today -Focus on what teacher saw in dashboard: weaknesses and strengths -Focus on Distribute Property and Combine Like Terms where most did not get to -Most reached Level 5 but not all -Focus on working with things they have never done before (Part B)	
LC9	1	Add/Subtract Constant From Both Sides	$x+8=-15$	No		
LC10	2	Divide Both Sides By The Variable Coefficient	$3x = 24$	No		
--	3	Distributive Property/Combine Like Terms		Yes		
LC8	a	(Stus started to fall off at the Distributive Property)	$5(x+4)=40$	Yes	-Focus on Distribution -Focus on two step equations: do add/subtract -Focus on canceling and simplification -Focus on checking answer	Teacher discusses with students, Student 7, 2, 5, 24
			$3=7(4-2u)-6u$	Yes	-Focus on Distributive Property and Combine Like Terms -Focus on distributing the 7 -Focus on divide and the other steps	Teacher discusses with students, Student 14, 11, 4, 16
				Not planned	-Focus on not distributing to the other side -Focus not distributing to the 6u term	
LC6	b	(This has that combine in it)	$3(1+4n)-2(5n-3)=25$	Yes	-Focus on splitting stus in groups -Focus on distribution of both parenthesis -Focus on Combine Like Terms, add/subtract, divide -Focus on checking solution	Teacher discusses with students, students work in groups, Student 11
				Not planned	-Focus on distributing the negative -2 with the other negative -3 as it is tricky	

Class Session. Moving to the next link in the causal chain (link 4 Figure 1), we counted how many of the statements in the lesson plan that were based on information from Luna, actually made it into the class session. For each class session, we analyzed the joint set of all notes taken during the sessions by all coders. We created tables to compare the lesson plan with the class session (Table 3). Next to each statement of the lesson plan, columns were added to show (1) whether teachers covered the planned statement in class, (2) a summarized description of what they discussed, and (3) who was involved in the discussion during the class session. The categories under the column Covered indicate whether teachers covered that statement in class (*Yes/No/Not planned*, with the latter code meaning the teacher did something they did not plan for or did not say they were planning for).

3.4 RQ2: Does Teacher's Use of the Dashboard Help Students Learn Better?

We studied whether students in the experimental condition, where teachers used Luna to prepare for the class session, had higher learning gains attributable to the class session, compared to the control condition. We consider the learning gains from pre- to post-test. (These gains can be attributed to the class session led by the teacher, since there were no other learning activities in between the pre-test and post-test.) We had analyzable data for 242 students (students who missed the pre-test, class session or post-test were removed from the analysis). Seven independent graders and the first author graded the tests. Fleiss's Kappa was 0.98. The grading schema gave full credit for correct statements and no credit for incorrect statements.

4 Results

4.1 RQ1: How Does the Dashboard Affect Teachers?

Teacher's updated knowledge. Across 5 teachers in 8 experimental condition classes, we recorded on average 12.6 statements per class that were evidence of the dashboard affecting what teachers knew about their students (Updated Knowledge in Table 4). (We will refer to the statements learned from Luna as "learned statements.") There were slightly more such statements at the class level compared to the individual level (7.1 statements per class at the class level versus 5.5 statements per class at the individual level). Teachers seemed surprised more often by information at the individual level (on average 1.4 statements per class) than at the class level (on average 0.38 statements per class). Further, out of the 12.6 statements on average that provide evidence that teachers learn from Luna, 34.7% relate to things that students are not doing well (19.8% at the class and 14.9% at the individual level), while 29.7% relate to things they are doing well (19.8% at the class and 9.9% at the individual level). Thus, Luna's information affected the teacher's knowledge about the class overall and individual students. Furthermore, these learned statements are about students doing well and not doing well with roughly equal frequency.

Lesson plan. Moving to the next link in the causal chain (Lesson Plan in Table 4), 44.6% of the learned statements get incorporated in the lesson plans (5.6 out of 12.6 statements per class learned from Luna). At the class level, teachers include in the lesson plans 33.3% of the learned statements, compared to 59% at the individual level. This finding suggests that Luna prompted change in teachers' lesson plans, both with respect to the class as a whole and to individual students, though more so with respect to the latter. In addition, teachers include an average of 3.1 statements per lesson plan pertaining to students not doing well (24.7% of all learned statements), namely, 1.9 (14.9%) at the class level and 1.3 (9.9%) at the individual level. By contrast, they include only 0.75 statements per class (5.9% of the learned statements) pertaining to students doing well (Figure 4)! As a different way of looking at this contrast, teachers include in their lesson plans 20% of the learned statements regarding students doing well, whereas they include 71.4% of the learned statements regarding students not doing well. Thus, the knowledge that teachers gain from Luna is

accounted for in various ways in their lesson plans, in particular knowledge about where students are struggling.

Table 4. Effect of the dashboard measured as average number of statements per class.

	Class Overall			Individual Students		
	Updated Knowledge	Lesson Plan	Class Session	Updated Knowledge	Lesson Plan	Class Session
(✔)	1	0.13	0.13	0.5	0.13	0.13
(+)	5.8	2	1.4	3.6	2.6	1
(!)	0.38	0.25	0.25	1.4	0.5	0.5
G	2.5	0.13	0	1.3	0.63	0.5
B	2.5	1.9	1.6	1.9	1.3	0.63
N	2.1	0.38	0.13	2.4	1.4	0.5
Total	**7.1**	**2.4**	**1.8**	**5.5**	**3.3**	**1.6**

We also made informal observations as to how the information teachers learned from Luna made it into their lesson plans. At the class level, in 6/8 classes where teachers prepared the control before the experimental classes, they used as a basis for the experimental classes the plan they prepared for the control ones, but changed and adapted it based on Luna's information. For example, they planned to discuss specific topics students were having trouble with, or added and removed exercises or topics from the plan based on Luna's information. One teacher, who prepared for the experimental before the control class, based the lesson plan for the former entirely on the dashboard, focusing on discussing errors the class was having with example exercises Luna provided for each error. In addition, based on Luna's information, in 1/8 classes the teacher decided not to cover a topic because the class had mastered it, while another teacher planned what topics to cover for the rest of the week, after the class session. At the individual level, in 3/8 classes teachers planned to work one-to-one, during or after class, with students who were not doing well as shown by Luna, while in 2/8 classes one teacher decided they did not need to spend time with individual students, who despite initially not doing well according to Luna, had fixed the problems they had, also according to Luna. In 2/8 classes, teachers adapted a worksheet they planned to give students based on the information in Luna. And lastly, somewhat to our surprise, in 2/8 classes one teacher assigned students to work in groups during the class session, with group composition based on students' progress as shown by Luna. In conclusion, there is a variety of ways in which teachers incorporate in their lesson plans knowledge they gain from Luna both at the class and individual level.

Class session. Moving down the causal chain, teachers implement in the class session 60% of those planned statements (Figure 4), which is 26.7% of the ones they learned from Luna (13.9% at the class and 12.9% at the individual level). Furthermore, 17.8% of the learned statements about students not doing well make it to the class session (12.9% at the class and 5% at the student level), as opposed to 4% of the ones about students doing well. Thus, the knowledge teachers gain from Luna that makes it to the lesson plan also gets accounted for and reaches students in the class session.
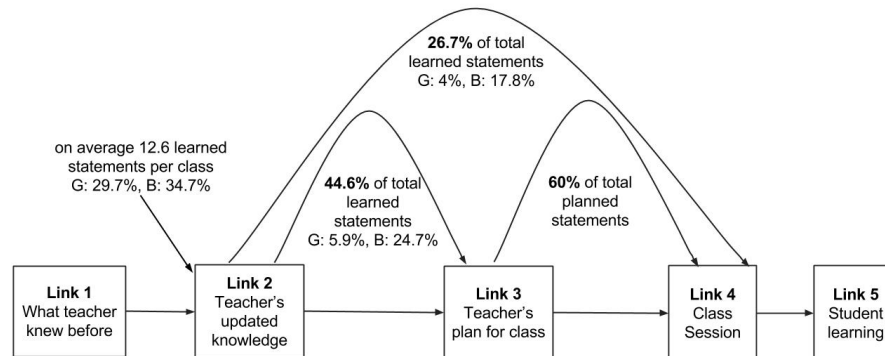
Fig 4. How the information from the dashboard traveled down the causal chain. The percentages on the arrows are percentages of the total number of statements teachers learned from Luna. "G" and "B" refer to statements about students doing well and not so well, respectively.

4.2 RQ2: Does Teacher's Use of the Dashboard Help Students Learn Better?

To test for knowledge differences between the conditions right before the class session, we ran a Welch Two Sample t-test on the pre-test data to compare the means of the control condition (M = 5.48, SD = 2.89) and experimental conditions (M = 4.53, SD = 3.23). We found that, in spite of our efforts to create balanced conditions, students in the control condition had a significantly higher pre-test mean than those in the experimental condition (t = 2.3908, df = 236.31, p = 0.0176). We used a hierarchical linear model (HLM [17]) with three nested levels to compare the gains from pre- to post-test (which can be attributed to the class session, with condition differences attributable to the dashboard). In the model, students (level 1) were nested within classes (level 2) which were nested within teachers (level 3). We included the condition as a fixed effect, and the difference between post- and pre-test as the dependent variable. There was no significant difference between the conditions in learning gains (t = -1.620, df = 240, p = 0.1065).

5 Discussion and Conclusions

We examine and trace the influence of a dashboard on teachers' knowledge of their students, their lesson plans and execution of these plans, and ultimately on student learning; these influences are summarized in a "causal chain" that guides our analysis. To the best of our knowledge, the use of this causal chain, to trace the effects of a dashboard for an ITS on teacher practices and student learning, is a methodological innovation in dashboard research. We note that this causal chain is not specific to ITSs or to the particular dashboard used. Further, to the best of our knowledge, the current study is one of the first that tries to measure *student learning gains* due to the teacher's use of a dashboard in a classroom setting [18, 22], with the exception of [7].

Our results show that the dashboard affects teachers at all the links in the causal chain. First, teachers update their knowledge with an average 12.6 statements per

class (Figure 4). In turn, the teachers' updated knowledge helps them to adapt or change their lesson plan. Teachers incorporate 44.6% of the statements they learned from the dashboard in their lesson plans, which suggests that Luna provided useful information to teachers on their students' performance in the ITS. Furthermore, teachers implement in the class session 60% of the planned statements, which is 26.7% of the statements they learned from the dashboard (Figure 4). This is a substantial portion, even if as we move down the causal chain, the number of statements that can be attributed to the dashboard decreases at every link. Perhaps that kind of "dilution" of influence, as we look at causal effects further removed from what teachers gleaned directly from the dashboard, is not surprising, although we believe our study is the first to document this phenomenon regarding dashboards.

In addition, we found teachers attend mostly to information from Luna that shows their students are not doing well in certain aspects of equation solving, as opposed to information about doing well. This perhaps is not surprising in and of itself but it suggests that the dashboard presents information that teachers do not have. Furthermore, although teachers learn almost the same number of statements for both the class overall and individual students who are not doing well, more statements related to the class, rather then individuals, get accounted for in the class session. Lastly, contrary to our expectation, we did not find that Luna influenced student learning. Generally, we can conclude that the dashboard's information, about skill mastery, occurrence of errors and student progress in an ITS, at the class and individual level, is helpful to teachers as they prepare for a class session, even if more is needed to demonstrate an improvement in student learning.

There are reasons to think that a fully designed dashboard, used over an extended period of time, could be even more influential than we found in the current study. First, as mentioned, at the time of the study, Luna was a high-fidelity dashboard prototype with some interactivity. A complete dashboard might provide more opportunities for teachers to look at more detailed information about their students' learning or might provide an option to project the dashboard in front of the class (cf. [7]). Second, the planning sessions were only 20 minutes total (for creating two or three lesson plans), which in retrospect was not enough time for teachers to fully digest Luna's information and plan what to cover in class. The class session was only 40-minutes, which restricted how much teachers planned for and covered. These time limitations could explain why teachers only planned for part of the information they learned from Luna and why fewer statements made it into the class session. Third, students took the post-test either right after the class session or the day after. Thus, they had no time to practice what teachers covered in the class session. Fourth, the dashboard was a new technology for teachers; the study gave them only limited time to become familiar with it, not enough to integrate it into their daily routines. In addition, only 2 out of the 5 teachers had previously worked with an ITS. When Luna is fully developed, with more opportunities for teachers to look at detailed information, and when used for longer periods of time, it could potentially help teachers bring more information from the dashboard into the class session, and ultimately help their students achieve higher learning gains.

In sum, the results of our study indicate that a dashboard with analytics from an ITS, based primarily on its student modeling methods, can be helpful to teachers. We

found that the dashboard's information affects the teacher's knowledge, lesson plans, and what they cover in the class session. In particular, the teachers paid much attention to their students' struggles. In our previous work [25] we found that teachers can have surprisingly detailed knowledge about their students; it was therefore not obvious that the dashboard would tell them much that they didn't already know. However, our study shows that even though teachers generally know their classes well, a dashboard with analytics from an ITS can still help them know more about their students, and can influence their lesson plans and lesson.

6 References

1. Aleven V., McLaren B.M., Sewall J., van Velsen M., et al.: Example-Tracing tutors: Intelligent tutor development for non-programmers. International Journal of Artificial Intelligence in Education 26, 224-269, (2016)
2. Aleven, V., Xhakaj, F., Holstein, K., & McLaren, B. M. Developing a Teacher Dashboard For Use with Intelligent Tutoring Systems. In the Proceedings of the 4th International Workshop on Teaching Analytics, IWTA 2016 at the 11th European Conference On Technology Enhanced Learning, EC-TEL 2016. Lyon, France, September 13-16, 2016. (2016)
3. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. Cognitive tutors: Lessons learned. The Journal of the Learning Sciences, 4(2), 167-207, (1995)
4. Arroyo, I., Woolf, B. P., Burleson, W., Muldner, K., Rai, D., & Tai, M. A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. International Journal of Artificial Intelligence in Education, 24(4), 387-426, (2014)
5. Hanington, B., & Martin, B. Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers. (2012)
6. Holstein, K., Xhakaj, F., Aleven, V., & McLaren, B. M. Luna: A Dashboard for Teachers Using Intelligent Tutoring Systems. In Proceedings of the 4th International Workshop on Teaching Analytics, IWTA 2016 at the 11th European Conference On Technology Enhanced Learning, EC-TEL 2016. Lyon, France, September 13-16, 2016. (2016)
7. Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., & Goldstein, D. S. Estimating the Effect of Web-Based Homework. In Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.), The Artificial Intelligence in Education Conference, pp. 824–827. Springer-Verlag (2013)
8. Kulik, J. A., & Fletcher, J. D. Effectiveness of Intelligent Tutoring Systems: A meta-analytic review. Review of Educational Research, 86(1), 42–78, (2016)
9. Long, Y. & Aleven, V. Mastery-Oriented shared student/system control over problem selection in a linear equation tutor. In Micarelli A., Stamper J., Panourgia K. (eds.), Proceedings of the 13th International Conference on Intelligent Tutoring Systems, ITS 2016, pp. 90-100. Springer International Publishing (2016)
10. Lovett, M., Meyer, O., Thille, & C. The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. Journal of Interactive Media in Education (2008)

11. Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology, 106(4), 901, (2014)

12. Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. The LATUX Workflow: Designing and Deploying Awareness Tools in Technology-enabled Learning Settings. Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, LAK 2015, pp. 1–10, (2015)

13. Martinez-Maldonado, R., Yacef, K., Kay, J., & Schwendimann, B. An interactive teacher's dashboard for monitoring multiple groups in a multi-tabletop learning environment. In Proceedings of the International Conference on Intelligent Tutoring Systems 2012, ITS 2012, pp. 482-492. Berlin, Springer (2012)

14. Mavrikis, M., Gutierrez-Santos, S., & Poulovassilis, A. Design and Evaluation of Teacher Assistance Tools for Exploratory Learning Environments. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, LAK 2016, pp. 168–172. New York, NY, USA: ACM. (2016)

15. Mazza, R., & Dimitrova, V. CourseVis: a graphical student monitoring tool for supporting instructors in web-based distance courses. International Journal of Human Computer Studies. 65(2), 125-139, (2007)

16. McLaren, B. M., Scheuer, O., & Miksatko, J. Supporting collaborative learning and eDiscussions using artificial intelligence techniques. International Journal of Artificial Intelligence in Education. 20(1), 1-46, (2010)

17. Raudenbush, S. W., Bryk, A. S.: Hierarchical Linear Models: Applications and Data Analysis Methods, Newbury Park, CA: Sage Publications (2002)

18. Schwendimann, B.A., Rodríguez-Triana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A., Gillet, D. & Dillenbourg, P. Understanding learning at a glance: An overview of learning dashboard studies. In D. Gasevic, G. Lynch, S. Dawson, H. Drachsler, & C. P. Rose, Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, LAK 2016, pp. 532-533. New York: ACM. (2016)

19. Steenbergen-Hu, S., & Cooper, H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. Journal of Educational Psychology. 106(2), 331-347, (2014)

20. van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. Supporting teachers in guiding collaborating students: effects of learning analytics in CSCL. Computers & Education. 79, 28–39, (2014)

21. VanLehn, K. The behavior of tutoring systems. International Journal of Artificial Intelligence in Education. 16(3), 227-265, (2006)

22. Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. Learning dashboards: An overview and future research opportunities. Personal and Ubiquitous Computing, 18(6), 1499–1514, (2014)

23. Waalkens, M., Aleven, V., & Taatgen, N. Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. Computers & Education. 60, 159-171, (2013)

24. Woolf, B. P. Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kauffman (2009)

25. Xhakaj, F., Aleven, V., & McLaren, B. M. How teachers use data to help students learn: Contextual inquiry for the design of a dashboard. In Verbert, K., Sharples, M., Klobučar, T. (eds.), Proceedings of the 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, pp. 340-354. Springer International Publishing Switzerland (2016)