Implementation of an Intelligent Tutoring System for Online Homework Support in an Efficacy Trial

Mingyu Feng[1], Jeremy Roschelle[1], Neil Heffernan[2], Janet Fairman[3], and Robert Murphy[1]

[1] SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025 USA
{mingyu.feng,jeremy.roschelle,robert.murphy}@sri.com
[2] Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609 USA
nth@wpi.edu
[3] University of Maine, Orono, ME 04469, USA
janet.fairman@maine.edu

Abstract. Much research has been done on the development of an intelligent tutoring system (ITS), and small empirical studies have demonstrated the effectiveness of ITS at promoting student learning. However, large-scale implementation of ITS in school settings has not been researched thoroughly. In this paper, we describe an ongoing randomized controlled trial (RCT) to evaluate the efficacy of a web-based tutoring system—the ASSISTments—as support for homework. The program is used in 46 middle schools in the state of Maine, to provide immediate feedback to students, and to provide reports to teachers to support homework review and instruction adaptation. We describe the challenges for the RCT, approaches used to understand implementation of the system, and findings on how the system is being used.

Keywords: efficacy, implementation, Intelligent Tutoring System, homework.

1 Introduction

The field of intelligent tutoring systems (ITS) has a long history and many studies have been conducted to show the effectiveness of ITS at improving student learning (e.g., Anderson et al., 1995; Koedinger et al., 1997; VanLehn et al., 2005). Recently, VanLehn (2011) claims that ITS can be nearly as effective as human tutors. Given the promising results found, efforts have been made to introducing ITSs into schools in order to help students learn more effectively (e.g., Koedinger et al., 1997; Arroyo et al., 2009). Most of these research studies have been at a relatively smaller scale within one school, or one school district in short durations. While these studies have the advantages of being more cost-effective and able to show the results quickly, factors such as varieties in school settings, implementation fidelity, counterfactuals, user support, and user-learning curves are typically not well studied and understood. After evaluating the Cognitive Tutors Algebra I (CTAI) curriculum, one of the most well developed ITSs, in a wide variety of middle schools and high schools in seven states for 2 years, Pane et al. (2013) reported there were no effects in the first year of

S. Trausan-Matu et al. (Eds.): ITS 2014, LNCS 8474, pp. 561–566, 2014.

implementation but strong evidence in support of a positive effect in the second year. One possible reason is that the teachers improved their implementation of CTAI or recommended instructional practices after a “warm-up” year of using it (Karam et al., submitted).

Homework is a well-established practice in schools, and the research knowledge base for the effectiveness of homework is also well established (Cooper et al., 2006). Yet, without explicit interventions, homework has been commonly underutilized for improving teaching and learning. Educational technologies have gained popularity in schools (e.g., Khan Academy, DreamBox, IXL.com), but not at home. Most of the computer programs for homework are for college-level populations (e.g., WebAssign, Mastering Physics, OWL) and many have been shown to have positive effects on learning (e.g., Dufresne, et al., 2002; VanLehn et al., 2005). However, there are few rigorous independent studies of the efficacy of online homework in K-12 settings.

The Maine Learning Technology Initiative has implemented one-to-one computing and supplied every seventh-grade student and their teachers with laptop computers, and most middle schools allow students to take their laptops home. In a randomized controlled efficacy trial (RCT), we are investigating whether a web-based ITS as a homework intervention can leverage Maine’s one-to-one laptop program to help improve student outcomes in mathematics as measured by a standardized test. Our focus in this paper is on the implementation of the ITS as an online homework support.

2 Background: The ASSISTments System

ASSISTments (www.assistments.org) is a web-based tutoring system that provides “formative assessments that assist.” Teachers choose (or add) homework items in ASSISTments and students can complete their homework online. As students do homework in ASSISTments, they receive immediate feedback on the correctness of their answers. Some problem types also provide hints on how to improve their answers, or help decompose multistep problems into parts (see Fig. 1). Teachers may choose to assign problem sets called “skill builders” that are organized to promote mastery learning (Anderson, 2000). Teachers also receive reports on their students’ homework and can use this information to organize more targeted homework reviews, to assign specific follow-up work to particular students, and to more generally adapt or differentiate their teaching.

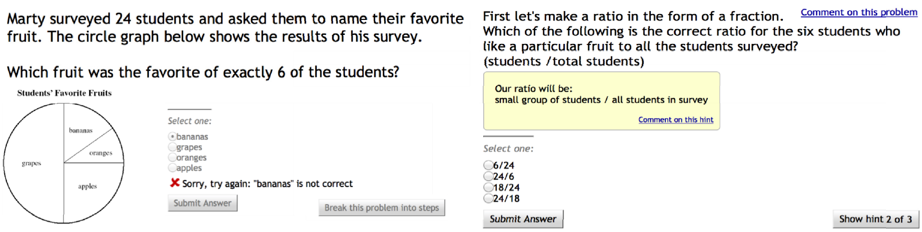

Fig. 1. Screenshots of a seventh-grade problem in ASSISTments that provides correctness feedback and break the problem into steps (left) and the first substep with a hint message (right)

Prior research also has established the promise of ASSISTments for improving student outcomes in middle school mathematics through homework support (Mendicino et al., 2009; Singh et al., 2011; Kelly et al., 2013). While the findings from these studies are encouraging, they only examined tightly controlled implementation of ASSISTments in a few schools over short durations. An investigation was not done regarding the factors that may hinder or facilitate the implementation of the intervention, which is critical for introducing the system to schools at scale.

3 Method

3.1 The Research Design

The study is an independent RCT involving 46 public schools from two cohorts, involving 114 teachers and more than 2,500 students in Maine, with schools randomly assigned to either treatment or control (i.e. "business as usual") conditions. The intervention is implemented in Grade 7 math classrooms in treatment schools over 2 consecutive years (academic years 2012–13 and 2013–14 for Cohort 1 schools and 2013–14 and 2014–15 for Cohort 2 schools). In the treatment condition, teachers receive professional development (PD) and use ASSISTments in the first year to become proficient with the system, and then teachers use ASSISTments with a new cohort of students in the second year when student outcomes are measured.

During the study, teachers in the treatment group are expected to assign approximately 25 minutes of homework in ASSISTments for a minimum of three nights per week, in order to take full advantage of the ITS. Homework assignments are expected to be a mixture of different problem types, including mastery learning problems, reassessment problems that are automatically assigned by the system, and textbook problems. Teachers will receive performance reports early the next morning via email.

The ultimate research question for the study is "*Do students who use ASSISTments for homework learn more than students who do homework without ASSISTments?*" While we are not there yet to answer this question, we hope to address an exploratory question through the data collected in the first implementation year: "*What is the implementation compliance and how much is ASSISTments used by students and teachers on learning?*"

3.2 Collecting Data at Different Stages to Facilitate Implementation

Data collection activities in the first implementation year center on understanding implementation start-up issues and identifying areas of implementation that may require additional support from the developer during the second implementation year.

Before Intervention: Understand the Context and Collect Baseline Data. A good understanding of the context of an RCT and the baseline information of the participants is needed to judge the impact of the intervention and to ensure the successful implementation of the intervention. At the beginning of the study, we conducted a 30-minute interview with principals from each school to learn about existing homework policy, data use, and other initiatives in participating schools. A pre-intervention teacher survey was administered to collect initial data about their current homework assigning, grading, and reviewing practices; formative assessment and differentiated instruction practices; and how technologies have been utilized to support homework.

During Intervention: Monitor Implementation Fidelity. In contrast to an effectiveness trial, the goal in an efficacy trial is to determine whether an innovation has a beneficial effect in *best-case* implementations. Therefore, it is fair game to monitor and adjust implementation of the innovation. ASSISTments automatically records detailed, time-stamped data of each student and teacher usage (i.e., "the click stream"). Analyzing such data allows us to assess the extent to which students are using the system to complete homework and the extent to which teachers are assigning problems and monitoring students' nightly homework performance. The design of candidate analytics can be guided both by the categories of implementation fidelity (e.g., adherence, exposure, quality of delivery, uptake; Cordray, 2008) and by the pathways in the theory of change. By doing so, a portrait of implementation is presented to the developer team, so that they can ponder: *Is this the quality of implementation we expected as creators of the intervention? What actions can we take that might bring implementation up to our desired levels*?

Halfway through Intervention: Capture Factors That Hinder Implementation. Near the end of the first implementation year, the team conducted face-to-face interviews with a random sample of the teachers to learn about (a) factors that influenced decisions related to homework assignments, (b) teachers' perspective on the impact of ASSISTments, (c) changes in teachers' review routines and instruction strategies, (d) challenges and usability of ASSISTments, and suggestions for improvement.

During Second Year Implementation: Establish Contrast with Counterfactuals. To attribute cause and effect between interventions and outcomes, one critical task of an RCT is to compare the implementation of the intervention with counterfactuals. After a "warm-up," routines have been set up to implement the intervention, and thus the focus of data collection may shift to establish contrast between the two experimental groups. Classroom observation is a powerful tool to capture teachers' practices and their interactions with students. We developed a classroom observation protocol to characterize teachers' reviews of homework and their efforts to adapt instruction. To better understand the motivation behind instruction adaptation, observers follow up with a brief interview.

4 Findings

Below we report preliminary findings from data collected from the first year of teachers and students' usage of ASSISTments.

The **principal interviews** revealed that in general homework is required and assigned almost nightly. This confirms that homework, despite all the controversial discussion regarding its influence on learning (Kohn, 2006), remains a major practice at schools. Teacher support was brought up as one of concerns as there were many demands on teacher's time (e.g., Common Core curriculum integration, meeting AYP goals, etc.) and a new intervention just added to these. We also learned access to the Internet at home is a concern in many schools. These perspectives were brought back to the PD specialist and the system engineers of ASSISTments. A teacher support plan was then adapted to make it more on-demand and ongoing, to better align with school PD community timelines and topics. An off-line version of ASSISTments was developed to ensure accessibility for all students during the study.

Teachers' responses to the **pre-intervention survey** revealed that teacher's general homework assignment practices align with the specified use model. Notably, even though Maine's laptop initiative has put laptop computers in the hands of every middle school student and teacher ever since 2002, we were surprised that no teachers chose "on laptop" when being asked, "*In what formats do your students usually do their homework?*" Among all of the 31 items in the survey, no significant differences have been detected between responses in the two different conditions.

Compared to self-report or observations, we found using **analytics of system logs** to monitor implementation fidelity is objective, and has lower cost and faster turnaround time. A first useful analytic was how often teachers made assignments with ASSISTments. We found that across 3 months, on average, most teachers assigned homework in ASSISTments 1–2 days in a week with only one teacher meeting the expectation of three assignments per week. Homework completion rates were around 75% and average minutes spent doing homework was 15 minutes. Both values were approximately as expected. A key "uptake" analytic was whether teachers were opening ASSISTments reports as a necessary prelude to adaptive teaching. The ASSISTments trainer was very surprised at the particular teachers who were not opening reports. These findings led to concrete plans of which teachers to follow up in the next round of school visits, what types of behaviors to target during coaching, and a change of the agenda items of the "best practices" workshop.

Although homework could provide data for adjustment of instruction, it is very time-consuming for teachers to aggregate and organize paper-based homework to scan for insights. Therefore, the **teacher interview** focused on the impact of ASSIST-ments reports on homework review. The biggest change reported by the majority of the 12 interviewees is that they can target on the problematic areas identified by the reports. The conversation shifts from checking correctness of every problem to "why" answers were wrong and the process of doing math. The homework review time reduces from 30 minutes to 15 minutes, as one teacher reported. The reports informed their planning and sometimes they had to change their plans when the report suggested students were not ready to move along. Teachers felt students were more engaged in the homework discussion because the discussion was more in time and on target. Based on the feedback from interviews, improvements were made regarding the usability of ASSISTments interface, accessibility of reports, and individual coaching.

5 Conclusion

In this paper, we present approaches used in an efficacy trial being conducted in 46 middle schools in Maine to collect data to understand the implementation of an ITS and thus better interpret the impact of the intervention on student learning. Overall, our recommendation is that researchers who are conducting RCTs to evaluate effectiveness of ITSs or other technology-based interventions in schools should focus on implementation and use different approaches to collect data at different stages to compare the implementation against a program logic model. This can lead to better control of the expected contrast between conditions, which in turn can improve the quality of the research. The implementation data also provides a unique opportunity for researchers to learn about the value that teachers and students find from the intervention, which is often non-detectable from a 30-item standardized test given at the end of the year. Research methods presented in this paper can be informative to later studies that aim at implementing ITS interventions at scale to a large population.

Acknowledgments. The research reported here was supported by the Institute of Educational Sciences, U.S. Department of Education, through Grant R305A120125 to SRI. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

1. 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)
2. Anderson, J.R.: Learning and memory: An integrated approach, 2nd edn. John Wiley and Sons, Inc., New York (2000)
3. Arroyo, I., Cooper, D., Burleson, W., Woolf, B.P., Muldner, K., Christopherson, R.: Emotion sensors go to school, Conference on Artificial Intelligence in Education (2009)
4. Cordray, D.S.: Fidelity of intervention implementation. Paper presented at the IES, Research Conference (2008)
5. Cooper, H., Robinson, J., Patall, E.: Does homework improve academic achievement? A synthesis of research, 1987-2003. Review of Educational Research 76, 1–62 (2006)
6. Dufresne, R., Mestre, J., Hart, D.M., Rath, K.A.: The effect of web-based homework on test performance in large enrollment introductory physics courses. Journal of Computers in Mathematics and Science Teaching 213, 229–251 (2002)
7. Karam, R., Pane, J.F., Griffin, B.A., Slaughter, M.E.: Evaluating Cognitive Tutor Algebra I curricula at scale: Focus on implementation (submitted)
8. Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., Soffer Goldstein, D.: Estimating the effect of web-based homework. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 824–827. Springer, Heidelberg (2013)
9. Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M.A.: Intelligent tutoring goes to school in the big city. Journal of Artificial Intelligence in Education 8, 30–43 (1997)
10. Kohn, A.: The homework myth: Why our kids get too much of a bad thing. Da Capo Press, Cambridge (2006)
11. Mendicino, M., Razzaq, L., Heffernan, N.T.: Comparison of traditional homework with computer supported homework: Improving learning from homework using intelligent tutoring systems. Journal of Research on Technology in Education (JRTE) 41(3), 331–359 (2009)
12. Pane, J.F., Griffin, B.A., McGaffrey, D.F., Karam, R.: Effectiveness of Cognitive Tutor Algebra I at Scale. Working paper (2013)
13. Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N., Razzaq, L., Dailey, M.: Improving K-12 homework with computers. In: Proceedings of the Artificial Intelligence in Education Conference, Auckland, New Zealand (2011)
14. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist 46(4), 197–221 (2011)
15. VanLehn, K., Lynch, C., Schulze, K., Shapiro, J.A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., Wintersgill, M.: The Andes physics tutoring system: Lessons learned. International Journal of Artificial Intelligence and Education 15(3), 1–47 (2005)