The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching

Neil T. Heffernan · Cristina Lindquist Heffernan

Abstract The ASSISTments project is an ecosystem of a few hundred teachers, a platform, and researchers working together. Development professionals help train teachers and get teachers to participate in studies. The platform and these teachers help researchers (sometimes explicitly and sometimes implicitly) simply by using content the teacher selects. The platform, hosted by Worcester Polytechnic Institute, allows teachers to write individual ASSISTments (composed of questions with answers and associated hints, solutions, web-based videos, etc.) or to use pre-built ASSISTments, bundle them together in a problem set, and assign these to students. The system gives immediate feedback to students while they are working and provides student-level data to teachers on any assignment. The word "ASSISTments" blends tutoring "assistance" with "assessment" reporting to teachers and students. While originally focused on mathematics, the platform now has content from many other subjects (e.g., science, English, Statistics, etc.). Due to the large library of mathematics content, however, it is mostly used by math teachers. Over 50,000 students used ASSISTments last school year (2013–4) and this number has been doubling each year for the last 8 years. The platform allows any user, mostly researchers, to create randomized controlled trials in the content, which has helped us use the tool in over 18 published and an equal number of unpublished studies. The data collected by the system has also been used in a few dozen peer-reviewed data mining publications. This paper will not seek to review these publications, but instead we will share why ASSISTments has been successful and what lessons were learned along the way. The first lesson learned was to build a platform for learning sciences, not a product that focused on a math topic. That is, ASSISTments is a tool, not a curriculum. A second lesson learned is expressed by the mantra "Put the teacher in charge, not the computer." This second lesson is about building a flexible system that allows teachers to use the tool in concert with the classroom routine. Once teachers are using the tool they are more likely to want to participate in research studies. These lessons were born from the design decisions about

N. T. Heffernan (✉) · C. L. Heffernan
Computer Science Department, Learning Sciences and Technology Program, Worcester Polytechnic Institute, Worcester, MA, USA
e-mail: nth@wpi.edu

what the platform supports and does not support. In conclusion, goals for the future will be presented.

Introduction

The mission statement for ASSISTments is "To improve education through scientific research while not compromising student learning time." Students and teachers use ASSISTments as a learning tool, rarely thinking of the research we are doing. Research is done by mining the data collected and by conducting randomized controlled trials (RCTs) which are embedded almost invisibly in the system's content. The studies fit into the normal routine in the classroom and so neither form of research compromises the students' learning time. The teachers, students, and researchers are part of what we call the ASSISTments Ecosystem. All of these players are part of making this mission a reality.

The ASSISTments platform assists *while* it assesses. Students find out immediately if they had the wrong answer to a problem allowing them to try again right away. Teachers get assessment results in real time. They can use this information to plan their next lesson, bring attention to common misconceptions, and group students for remediation and acceleration. Last year 50,000 students made accounts. On average, 4,000 students use ASSISTments each weekday, about half of them using it for nightly homework, while the rest are using it in school. The authoring tool, designed so that teachers and researchers can write content, has been key to this success (Razzaq et al. 2009). The tool also supports researchers in creating randomized controlled experiments. About a dozen universities now are using ASSISTments to run studies of some sort. The tool has been used to do randomized controlled trials reported in over 18 peer-reviewed publications (Broderick et al. 2012; Heffernan et al. 2012a, b; Kehrer et al. 2013; Kelly et al. 2013a; Kelly et al. 2013b; Kim et al. 2009; Mendicino et al. 2009; Ostrow and Heffernan 2014a, b; Pardos et al. 2011; Razzaq et al. 2005, 2009; Razzaq et al. 2007; Razzaq et al. 2008; Razzaq and Heffernan 2006; 2009; 2010; Sao Pedro et al. 2009; Shrestha et al. 2009; Singh et al. 2011; Walonoski and Heffernan 2006). Those studies were generally about comparing different ways of giving feedback to students, and measuring learning on a posttest.

In addition to the above randomized controlled trials, ASSISTments has been used extensively as part of data mining research. Specifically, its data has been used in over 50 other peer-reviewed publications in the areas of predicting state test scores (e.g., Feng et al. 2009), using Bayesian networks (e.g., Pardos and Heffernan 2012), or using the platform to create detectors of students' emotional state (affect detectors; e.g., San Pedro et al. 2013). We do not review all of these studies here. Instead, this paper summarizes the history of ASSISTments, and what lessons learned are worth sharing. Van Lehn's (VanLehn et al. 2005) *Lesson Learned* paper points out that all systems involve thousands of little design decisions; not all of them can be shared. The two lessons most worth sharing with others in the field, that are not in any of our prior publications, are 1) to "build a platform" that can easily deploy a wide range of content and experiments and 2) to make the platform flexible so teachers will embrace it as their

own (putting the teacher in charge). By articulating how design decisions were made and how these decisions were influenced by research questions, others can learn from this experience.

ASSISTments and the Existing AIED Literature

The Artificial Intelligence in Education (AIED) community has applied AI in many different ways to try to improve education. For example, Lester et al. (2013) work focused on storytelling, natural language, and a concern for motivation and affect. Natural language processing (NLP) is clearly an AI area and many systems now try to do this (e.g., Person et al. 2001; Wang et al. 2011). Another use of AI is in trying to infer and respond to student emotion (e.g., Baker et al. 2010; du Boulay et al. 2010; D'Mello et al. 2010; Woolf et al. 2010). Still others have used speech recognition technology to try to better teach students (Chen et al. 2011; Johnson 2010). To create intelligent tutoring systems (ITS), some researchers have used AIED frameworks like the decision theoretic system (VanLehn et al. 2003) or reinforcement learning (Chi et al. 2011). While all of these are interesting applications of AI to education, this article comes out of the model-tracing work of building intelligent tutoring systems (Anderson et al. 1995; VanLehn et al. 2005). The model-tracing paradigm is one of two key paradigms that have been commonly used in creating intelligent tutoring systems. The model-tracing paradigm (Anderson et al. 1995) and the constraint-based paradigm (Mitrovic and Ohlsson 1999) have had major impacts in the community. Both have come out of perspectives of modeling knowledge and, in that sense, they have come out of an AI perspective. Anderson's ACT-R framework postulated that key aspects of human cognition can be modeled with production rules. Ohlsson presents an alternative theory of learning of constraints (Ohlsson 1992). Both of these theories have generated many systems that have posted impressive results (Koedinger et al. 1997; Suraweera and Mitrovic 2004). Under the model-tracing paradigm, as articulated in Anderson et al.'s (1995) *Cognitive Tutors: Lessons Learned* paper, the steps in creating a tutor are 1) "selecting a problem-solving interface," 2) "constructing a curriculum under the guidance of domain expert," 3) "designing a cognitive model for solving the problems," and 4) finally "building instruction around the cognitive model" in the form of hints and feedback messages for incorrect steps.

While model-tracing tutors have had impressive research results (e.g., Koedinger et al. 1997; VanLehn et al. 2005), ASSISTments has evolved, in some respects, to be the antithesis of an AIED system. There is not a strong "student model" as recommended by Woolf (2009), nor does the system do any deep reasoning that some might expect of an "intelligent" tutoring system. Sure, we have a tagging infrastructure that allows us, or teachers, to tag a question with a skill, but that skill information does not directly drive problem selection. In fact ASSISTments does not know how to solve the problem (contrary to Anderson et al.'s suggestion that you "build a cognitive model for solving the problems"). This will be discussed more in the model-tracing in the history section.

When ASSISTments was first created, it was positioned as an intelligent tutoring system, with behavior close to functionally equivalent to that of a model-tracing tutor (Koedinger et al. 1997). But now, looking back on a decade of work, things look

different. This will be covered in the next section on the history of ASSISTments where it will be shown how ASSISTments has changed and grown over time.

History of ASSISTments

ASSISTments was born from needs uncovered while we (the authors) were teaching 8th grade math. Teachers need to understand what students know and don't know - to understand their prerequisite knowledge. Tracking students' mastery of skills has traditionally been done on paper, but doing so is a burdensome task for teachers that takes time away from teaching or preparing to teach. Teachers face this sort of problem daily – how to glean information on what students understand and to track the skills they have mastered. How can we help these teachers and students?

Starting with *Ms. Lindquist*

We teamed up to learn about how human tutors worked. Tutoring sessions were videotaped to learn how human tutors interacted with students. From these videos was born an AIED system, *Ms. Lindquist*, that mimicked some of the behaviors of human tutors (Heffernan and Koedinger 2000). It became clear one behavior occurred frequently: the human tutor asked a series of questions to help the students build understanding. Some have called these series of questions "micro plans" (McArthur et al. 1990). The key thing was that the human tutor knew what to ask and when. We have adopted the term *scaffolding question* to describe what the tutor asks as a way to support the solving of the original problem. The scaffolding question might simply break the problem into sub-problems, or it might use some sort of analog (like the "concrete articulation" strategy used by humans reported in Heffernan 2001, p. 39). It is an important part of this story that *Ms. Lindquist* was one of the first ITS to use the internet to perform experiments on learning (Heffernan 2003; Heffernan and Croteau 2004). Others had put intelligent tutoring system on the web already (e.g., Brusilovsky et al. 1996) but no one, to our knowledge, used such a system to create published randomized controlled trials. All of this inspired us to take this idea of a platform for scientific research to scale for other topics.

ASSISTments as a Tool for Building Tutors

Ms. Lindquist was made using the model-tracing tools developed at Carnegie Mellon, the Tutor Development Kit (TDK; Anderson and Pelletier 1991). The TDK allows an author to write rules (in a language similar to Anderson's ACT-R rules; Anderson 1993) that would *model trace* student actions (i.e., the model could generate problem solving steps in the way a student solved problems). Anderson et al.'s (1995) *Lessons Learned* Principle Number 8 was to make concrete or "reify the problem solving structure" so an author would use tools to allow students to show their intermediate steps in problem solving and get support while doing so. The TDK also supported knowledge tracing (tracking skills the student knew). The TDK had been used very successfully in creating the Algebra Cognitive Tutor (™) program (Koedinger et al. 1997). While the Cognitive Tutor program began as a research project that eventually became a

commercial success (as a product of Carnegie Learning, Inc), it was very costly to build. The cost estimates for intelligent tutoring systems vary between 100 and 1,000 h of work for every hour of content a student will see (Anderson 1993, p. 254; Murray 1999). Not only was the amount of time high, the skill level required for authors was very high, and it was recommended that the programmers in the cognitive psychology and AI communities work together to create rule based programs. Koedinger and his colleagues hosted multiple summer workshops where researchers were invited from across the world to learn TDK. While we don't know the exact numbers, most summer school participants found the process very difficult. Most left and did not go on to build model-tracing tutors, though some other labs have adopted the work (i.e., Blessing et al. 2006; Lebeau et al. 2009). Certainly no other lab replicated the success of CMU in building a research supported and commercially viable model-tracing tutor using TDK. We think the reason for this is that the process was too costly in terms of skill levels and time.

Around 2000, we joined others in an effort to dramatically reduce the cost of creating such tutors. This work produced two projects: Cognitive Tutor Authoring Tools (CTAT, Aleven et al. 2009a, b), and one that would become ASSISTments authoring tool (Razzaq et al. 2009). (The ASSISTments system version did not allow as much functionality as the CTAT tutor did, such as supporting multiple pathways.) The main idea of CTAT was that by writing examples of problem solutions, the author could model a system that was behaviorally equivalent to a model-tracing tutor. CTAT was first conceived as a tool to make it easier to write rule-based tutors by first asking authors to provide examples of the problem-solving behavior that the tutor should help students learn. CTAT then supported the next step of writing rule based programs. The system would show authors which examples their rules were able to model (i.e., the examples served as test cases). This line of work eventually doubled the efficiency of rule writing processes (Aleven 2010; Aleven et al. 2006).

Eventually CTAT was used to mimic many of the key behaviors of model-tracing tutors without authors actually writing rules. These examples in CTAT provide what is an example-tracing tutor as opposed to a model-tracing tutor (aka, a rule-based tutor). Koedinger et al. (2004) reported on the time required to create several projects using example tracing tutors and reported an eight-fold increase in efficiency compared to the time it takes to create model-tracing tutors.

The next section will explain some of the modifications made to this tool to mass produce hundreds of example tracing tutors. The following four sections will cover some of the history of the platform and approaches to teacher training. The dates are not consecutive, as the four different elements do, in fact, overlap.

ASSISTments as a test preparation program (2003–2005)

In 2003, along with Ken Koedinger, we used the yearly push, in American middle school and high schools, to prepare students for high-stake state tests as a starting point for building a new tutor. We realized that there was a problem with giving students more assessments as a way to predict who needed test prep in schools. Students needed to stop everything in order to take a practice test; teachers had to take time out of their planning periods to grade and analyze it; students got feedback on their performance only after the teacher had had time to do so. A computerized tool that gave

ASSISTance as it collected assessMENT data would allow for the benefits of the assessment without wasting student or teacher time. This idea is a win-win because teachers and students would get the feedback they need to succeed (and would get it immediately) and researchers would get test subjects since many teachers were interested in practicing for the tests. Massachusetts, releases all of the Massachusetts Comprehensive Assessment System (MCAS) test items each year, providing us, and ASSISTments, with a large bank of problems to use.

In 2003, the ASSISTments project began. We wanted to build an online system where students would practice the released MCAS items, with tutoring on how to work out problems offered to students who got the problem wrong. This helped teachers predict which students would pass the MCAS and hopefully help those who might not pass, before the next exam took place.

Initially CTAT was used to create the tutor, but within the first year we abandoned CTAT and began to write code to build ASSISTments. In doing so, we totally abandoned the rule writing step as well as eliminating other CTAT steps such as the interface specification step. One thing that CTAT did not support was the easy construction of scaffolding questions. We felt this was an important component of a tutor and therefore built ASSISTments so it provides a structure for scaffolding questions each with its own answer and hints.

We also went back to our roots of working with teachers. We nurture our relationships with the math department heads at the three middle schools in Worcester, MA. These relationships allowed us to get expert advice on the tutoring and provided subjects to pilot the program. We would show the teachers a released item and then videotape them as they discussed the solution. In particular teachers were asked what questions they would ask a student who was having trouble solving the problem. The "Two Triangles" item in Fig. 1 was created by Mr. Paul King at Forest Grove Middle School. The question happened to have been the 19th question on the 2003 MCAS test. Mr. King asked four scaffolding questions. First he asked about the congruence, then the perimeter, then the equation solving, and finally the substitution. Figure 1 shows an example of those questions. (To play the item and see all the questions, go to http://goo.gl/xIxVim.)

Undergraduates at WPI wrote hints and bug messages (responses to common wrong answers) for each of the questions. Since the teachers had been involved with the creation of the problems, they were excited to pilot them with their students. They saw the benefit of giving immediate feedback and support to many individual students all at the same time. This was the beginning of ASSISTments being a collaborative or ecosystem. These teachers were invested in the project. They created hundreds of ASSISTments (a single question is called an ASSISTment as it involves not just the question but also the answer and the associated assistance). Each ASSISTment was hand-written, and we eventually made hundreds of them. Because the authoring tool provided structure to the creation of the tutoring, the scaffolding questions and their hints were quite simple to create. Figure 2 shows this authoring environment that allows the teacher to create an ASSISTment (represented as an HTML page with the ability to upload images, links or embedded videos). The stubs for the scaffolding questions can be seen to the right of the image, two of which show up in Fig. 1.

Within a year, we took over the job of writing the scaffolding questions with the support of undergraduate students at WPI (over 100 have worked on the project

Fig. 1 A screen showing a student getting "tutoring" to help the student figure out how to solve a question

creating content). Eventually we would grow out of just authoring MCAS questions, but that was how we started. We also bundled these ASSISTments into problem sets. A problem set is a grouping of ASSISTments a teacher can assign. The ASSISTments can be delivered in random or linear order. We will describe later how users can also manipulate problem sets to create randomized controlled trials for research. The teachers had limited time available, and we used our time with them to develop the teacher reporting and assigning interface. Because of the importance of the state tests the teachers took their own students to the school's computer lab to get immediate

Fig. 2 A view of the builder that allows the author to write questions with images, as well as answer with feedback messages. Note that it shows that of the 4,928 students who attempt this problem (and did not first ask for a hint), 852 students said 5 as their first response showing a common wrong answer

feedback on their students' performance and so they could provide support with problem solving.

They also received useful information on what needed to be re-taught. These piloting experiences were essential in designing the user interface.

Early on, we built a report for teachers that predicted students' state test scores from their performance, using the ASSISTments systems, on released test items. PhD students Feng and Pardos designed and carried out many experiments around predicting these scores (Feng and Heffernan 2010; Pardos and Heffernan 2012). We think these reports were not, in fact, all that useful to teachers as they already knew which students were doing well overall, though the existence of the report may have been useful for "selling" the project to school leaders.

From this basic beginning, the format of ASSISTments continued to improve and develop. Along with growth of the system itself came growth in the number of teachers using the system. We learned that aiming ITS authoring tools at teachers, not instructional designers or other kinds of curriculum developers, can be very successful if the tools are easy enough to use. In the next section, we will describe how we recruited and trained the teachers who use ASSISTments with their students.

Scaling up Outreach (2004–2008)

In the first 3 years we only worked with local schools. During that time, one of us was almost always present when the tutor was used with the students. The use of a computer lab was the only reasonable way for middle schools to access computers with students. Access to computers for homework and tablets in the classroom was very uncommon. Therefore, the teaching strategy for using ASSISTments was to go to the lab approximately once every 2 weeks to practice MCAS problems related to what they were studying. The students received tutoring while the teachers received reports. The teachers saw ASSISTments as a computer tool to aid in their work. We built experimental manipulations into some of the problem sets, so we could run studies (Razzaq et al. 2005). For instance, we compared two different ways to support the students if they did not get the question correct. Students and teachers never felt they were interrupted with a study.

We received a grant to train our graduate students in communication skills by going in to schools (NSF, 2008). This grant charged us with training our graduate students to learn communication skills by going into schools. Part of the grant was to recruit and train teachers to use ASSISTments and then pair up these expert teachers with the computer science graduate program. This grant also included mentoring the graduate students' relationship with the teachers and began a new tradition of training teachers in using ASSISTments as a tool for formative assessment. As mentioned, the MCAS had a big influence on the initial content we built, and this content was used because teachers cared passionately about preparing their students for the MCAS. By 2008, teachers were being asked repeatedly to use data to inform their instruction by administrators and policy makers. ASSISTments could be the tool they were looking for to help with the bookkeeping involved in getting and using data.

In order to recruit more teachers for this program, we invited teachers and administrators to observe teachers and students using ASSISTments. We called these in-school events *ASSISTments in Action*, where teachers and administrators could observe other teachers in the classroom using ASSISTments and imagine how they would do the same. We also had Friday workshops once a month where teachers would pay a fee to be trained on how to use this free tool. Tutorials were put up on the web to demonstrate how to use ASSISTments. We recruited teachers who were graduates of our trainings to teach classes offered through local education collaboratives on how to use ASSISTments for data driven instruction. We visited over 30 school districts and trained their teachers on site. We recognized that the more teachers used the system, the more opportunity for research we would have. The teachers and schools on the other hand obtained a useful tool to help them and their students.

Mastery Learning, Skill Builders and the Automatic Reassessment and Re-learning System (2007–2010)

Now, let us turn back to the history of the building of ASSISTments. In 2007, we had hundreds of ASSISTments covering all released MCAS items, but we did not have enough questions to automatically assign practice for each skill that a student might be struggling with.

We were interested in mastery learning and without a bank of questions, students would not get enough practice to master a skill. Mastery learning is a method to make sure that students master a topic or skill before going on to the next topic, by assigning problems on a given topic until a mastery criterion is reached. This idea is also used in grading students: they do not get points off if they get something wrong early on but later show they know it.

Mastery learning has been a part of the ITS field for a long time (e.g. Bloom 1984). ASSISTments has never tried to do a study comparing mastery learning with a non mastery learning approach. Later VanLehn (2011) argued that Bloom's claim probably over estimated that effect size; but VanLehn's meta-analysis did report that mastery learning was quite effective. Others have also shown mastery learning to be effective in computer tutors (e.g., Corbett 2001).

We needed an infrastructure for building more content so ASSISTments could be used daily instead of every few weeks. We built a way to generate more content, harking back to the TDK rule-based system, but using a simpler approach than AI Rules. The author of the question would create a template with variables instead of real numbers. We would then randomize the numbers in the questions and hints, as well as randomized simple cover story elements (e.g., {Name1} had {x} marbles, {Name2} had {y} cookies) to create a large number of similar problems. Other systems like the aforementioned TDK and CTAT did/do have similar features. We could then make ASSISTments more adaptive since we had so much more content. The adaptive feature we built was called Skill Builders. The idea was to keep giving the student questions until some proficiency threshold was reached (by default it was "three correct in a row" but could be manipulated by the teacher).

Over the next 3 years we taught groups of undergraduates in building this content. They would pick a topic, create the skill builder problem sets using the template system and then as part of their junior year research project they would run some form of randomized controlled trial. For example, the Pythagorean theorem skill builder problem set has a different templates for each of the common problem types like 1) find the hypotenuse given two sides 2) find a side given the hypotenuse and a leg and 3) find a side given you are told it's a right isosceles triangle and also told the hypotenuse. The skill builder works even if we have not tagged the questions with knowledge components, however we tried to tag them to match other reports work correctly. In the third example above it was the case that we did not bother to tag it with the "isosceles" knowledge component, even though it clearly is related. We acknowledge that there are some inaccuracies in our modeling, but we did not want to prevent teachers from getting access to good content just because Dr Heffernan has not solved the hard credit/ blame problem (i.e. if a student gets the third template wrong it might be because they don't understand "isosceles" triangles).

Also, when creating a problem set we would have to decide what percentage of problems to create from each template. Some problem types are more common than others. I want to point out that the stopping criterion for a skill builder was not based on the skill-tags of the items within but instead is just based on getting three right in a row (some student might get "lucky" and receive three of the same template, which is one of the reason, we will talk about shortly, we created ARRS).

We hired a middle school math teacher to work over the summer and create skill builder problem sets. His expertise in using the system and pedagogical knowledge
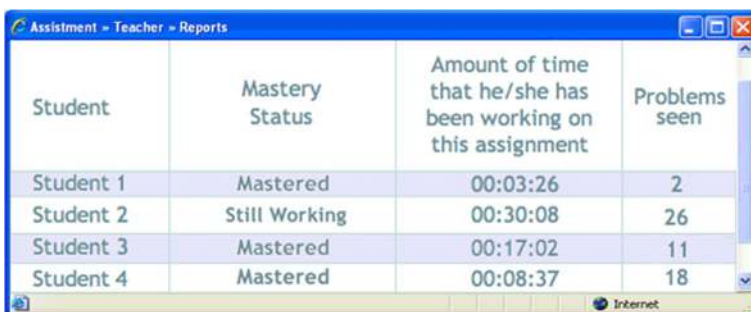
helps us run studies. Figure 3 shows a teacher reports showing that the first student was able to quickly master the skill. Student 3 also mastered the skill, but the report shows that this student struggled far more to reach mastery.

Once we had skill practice problem sets, the next logical step was to add spaced practice. Along with Joe Beck, we created the Automatic Reassessment and Relearning System (ARRS) to do just that (Li et al. 2013). Students are reassessed on a schedule of 7, 14, 28 and 60 days after initial mastery. By spacing out this practice we are looking to see if we can increase student learning. If they failed their reassessment they would be assigned a relearning activity and their spacing schedule would restart. We now have over 200 ASSISTments Certified Skill Builders and they are aligned to the new Common Core State Standards making them easy to find and very attractive to teachers. In a recent study we showed ARRS was effective at promoting retention (Heffernan et al. 2012a, b). In this study we compared Skill Builder practice by itself versus Skill Builder practice with ARRS, which gave extended reassessment and practice. In summary, skill builders combined and ARRS seems to provide a good bang for the buck. We will discuss in 4.1.2 how we can embed experiments into those Skill Builders.

Other Platform Features: 2008 - Present

ARRS and Skill Builders are an example of the type of adaptive features we built within the basic structure of ASSISTments. We have built other features and continue to so as needs present themselves.

Parent Notification While our focus has been on giving feedback to students and teachers, parents are an important part of the equation. We built the Parent Notification System that allows teachers to share the student data with parents. Broderick et al. (2012) reports on a randomized controlled trial where we randomly assigned students into conditions of "No parent access" or "Yes: parent access". In the "No access" condition the class proceeded as usual and teachers communicated with parents in their usual fashion, while in the experimental condition of "yes access" parents had accounts in ASSISTments which emailed them when their students failed to complete their work, and provided details on how they were doing. There was a reliable increase in parents' feeling of connectedness to what was happening in math class (via self reports) and an increased in homework completion rates.

Student	Mastery Status	Amount of time that he/she has been working on this assignment	Problems seen
Student 1	Mastered	00:03:26	2
Student 2	Still Working	00:30:08	26
Student 3	Mastered	00:17:02	11
Student 4	Mastered	00:08:37	18

Fig. 3 A Screenshot of a teacher's skill builder report showing student performance

Open-Response Questions We pride ourselves in freeing up teachers time by grading a variety of question types. From the beginning, ASSISTments has had item types such as multiple choice, choose all that apply, and rank, as well as an open box where students can write any number or its equivalent; for example, 2x +1 and 1+2x are both marked correct even if the author just gives the first one. But teachers asked us to add a field where students can just write an explanation. Once teachers grade these items they can send comments to their students and the grade will average in with other automatically graded problems. We then added an automated peer review system (Heffernan et al. 2012a, b) to help teachers and students review these open-ended responses.

Differential Instruction We have a feature that allows teacher to assign a new problem set depending on the grade on an original problem or problem set. Teachers see the average correct for the problem, say 67 % of the students go the problem correct. They can then assign one problem set to the 67 % of the students who got it right and another problem set to the 33 % who got it wrong.

EdRank A different feature we created is EdRank, a project named by Joe Beck, where we run randomized controlled experiments comparing the value of different web pages to teach different topics. For each skill we track, we searched the web to find various pages that purport to teach this topic. We measure student learning by treating the next question, tagged with the same skill, as a posttest item. Some very initial findings are reporting in Gong et al. (2012), but this project is just getting underway.

PLACEments Most recently we rolled out PLACEments (Whorton 2013), which like ARRS is connected to our Skill Builder content. PLACEments is a computer adaptive test that uses a prerequisite hierarchy of skills to decide what questions to give a student. This gives a teacher support on a type of differential instruction and support to students. A teacher initially selects what Common Core Standard to assess. The student is assessed on problems related to this standard, but for each error they make, the system will add, specifically for that student, assessment questions on that skill's prerequisites.

We think these features add value to the platform.

Different uses of ASSISTments have Increased Student Learning

In this section we will talk about some of the peer-reviewed results where ASSISTments has been used to increase learning. Note that we do not suggest that ASSISTments itself leads to better learning, as results depend entirely upon the content used within the platform. ASSISTments is merely a platform with a strict set of functions but that content created in it, and the order in which it is delivered to students, is what determines how much students learn. We will now describe some of the results that have shown student learning can be increased with the use of content within ASSISTments.

We have compared ASSISTments to a business-as-usual condition where students did homework on paper and did not receive feedback until school the next day. In Mendicino, Razzaq and Heffernan (2009), the control group received a worksheet with

problems, but the experimental condition received scaffolding and hints for each problem immediately after they got it wrong. For the control, teachers gave students feedback on homework in class. Posttest results show a reliable effect of condition and a meaningful effect size of 0.6. In follow-up studies we have found similar findings that showed ASSISTments mediated immediate feedback lead to better learning than the business as usual next day feedback (Kelly, et al. 2013b; Singh et al. 2011).

A second line of evidence that suggests ASSISTments is effective is provided by Koedinger et al. (2010) who analyzed schools in an urban city in New England that used ASSISTments. They compared performance of students attending schools in the same city system that did and did not use ASSISTments. They collected end of year state tests for all students, 2 years in row, so they could compute "growth" between the 2 year's tests. They reported reliable gain scores on state tests. A weakness with that study is that schools were not randomly assigned to treatment, but the results from Koedinger et al. (2010) and Mendicino et al. (2009) led the US Dept. of Education's Institute for Education Sciences to grant an efficacy trial of ASSISTments as a homework tool to SRI Incorporated to see if the platform and the content used in 7th grade math could lead to more learning. It uses school-level randomization, where half of the 48 schools are assigned to the control group and cannot use the system for 2 years. The study is being conducted in Maine, as all 7th and 8th grade students have computers due to a state initiative funded in 2002. Teachers did not have to alter class content because the team at WPI created all of the required "Book Work" questions. A progress report on this work was reported in Feng et al. (2014), with finally results not due out until 2016.

A third line of evidence that suggests ASSISTments is effective, was found by SRI. In 2012 the Gates Foundation funded 17 projects in the Next Generation Learning Challenges. They hired SRI to do an outside evaluation. ASSISTments was the only project that produced reliable increases of student learning (measured using the NWEA's MAPS standard test used by 5.1 million students) beyond what would be expected, as measured by comparing to a control group (Miller et al. 2013).

Finally, a fourth line of evidence suggesting efficacy was reported by Jim Pellegrino and colleagues. In this case, they were reporting on the effect of the Automatic Reassessment and Releaning System, and compared it to a condition that only gave practice at the introduction of the topic. In these studies the number of problems a student did was fixed, but in Skill Builders the student needs to keep going until they reach some initial proficiency on the topic (typically, 3 correct in a row). Dr. Pellegrino has documented that Skill Builders and ARRS increase student learning (Soffer, et al. 2014).

Now that we have described how ASSISTments has been used to show improved learning, we will next describe how researchers have already conducted RCTs within ASSISTments.

Lessons Learned

Our goal has been to build a platform to study what works in education. We never thought of ourselves as building the best content in the world, but we needed to build enough content to get teachers to start to use the system. So the point is that we needed

a platform with a community of teachers. But to get teachers to use adopt it, we needed to put them in charge. In the next section we will address first the "build a platform" lesson, followed by the "put the teacher in charge" lesson.

Lesson One: Build a Platform

There are many reasons to build a platform to support this work, but the most important reason is that unless we have a flexible easy to use platform it is hard to use it for science. Our advisory board member James Pellegrino, Professor at the University of Illinois Chicago and member of the National Academy of Education, had this to say about ASSISTments:

> "The combination of WPI's ASSISTments and the collaborative relationships with numerous schools and teachers provides an incredible testbed for designing and evaluating…instructional materials. There is no place else in the country where such a capability exists."

Even though we have been building ASSISTments since 2003 to perform randomized controlled trials, when he used this "testbed" metaphor, it helped us realize that our goal should be to help others use this platform to conduct good research using randomized controlled designs. In a recent article in *Science*, Koedinger et al. (2013) estimated that the science of instruction is faced with over 300 trillion options regarding how to design instruction. How are we as a scientific community going to explore this overwhelming set of options? We think our community of scientists, teachers, parents, and students, needs platforms that make it more efficient to test more of these options. ASSISTments could be one of these platforms.

There are two main ways RCTs are run in ASSISTments. The first is from the back end manipulating the features. This type is done in-house at WPI since it involves adding and manipulating feature in ASSISTments, adjusting the code of the ASSISTments system. The second type of RCTs is done by the users of the system, using the authoring tool (and its extension that allows for randomization called a "choose" conditions). They build an RCT into a problem set (see Figs. 4 and 5 showing the editors). Anyone can use this second type; they just need an account in ASSISTments, an idea, and us to expose their content to our subject pool.

Experiments Controlled from the Back End

An example of the first type of RCT, which we control from the back end, is within the ARRS system. Xiong and Beck (2014), both at WPI, reported on an experiment where they compared different spacing regimes. In such studies, students usually do not know they are getting something different from their classmates; they are simply assigned a reassessment test and do it. Some students see the test after 2 days while others wait 7 days.

Experiments Controlled by End Users

As an example of the second type of RCT that anyone can create using the ASSISTments authoring tools, Shrestha et al. (2009) did a study that compared worked
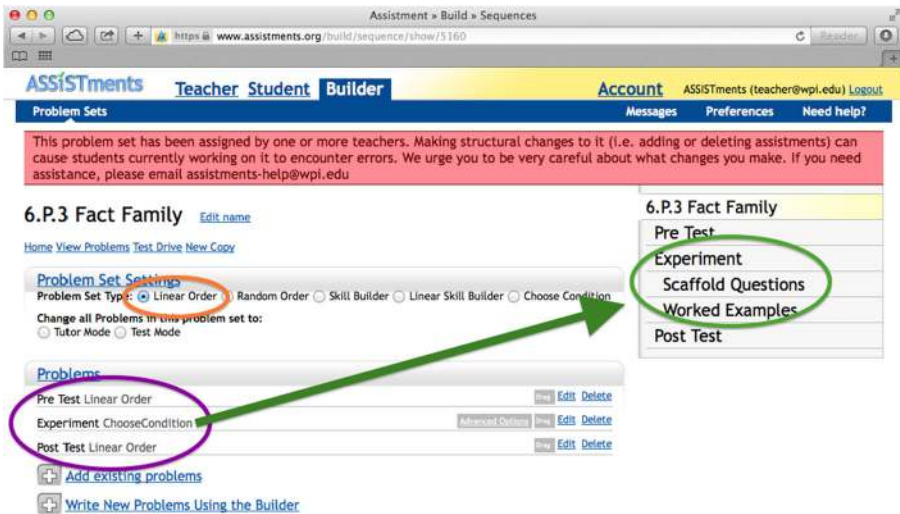
Fig. 4 ASSISTments authoring tool with which a researcher can create a problem set with embedded RCT. In this case the problem set is of type "Linear Order" (see orange oval) where students will do three sections in a row (see purple oval): first they will do a section called "Pre Test," followed by one called "Experiment," and then finally a section called "Post Test" (the names for sections are created by the experimenter and have no meaning except to help the experimenter to remember his design). The experimental section is of type "ChooseCondition," which means it will randomly pick one of the objects contained within it. The conditions in the section labeled as "Experiment" are called "Scaffold Questions" and "Worked Examples" (see the green oval)

examples with standard ASSISTments scaffolding (i.e., using scaffolding questions). Shrestha et al. were WPI undergrads; they used the tool to create 19 randomized controlled trials where students were randomly assigned to get one of two types of

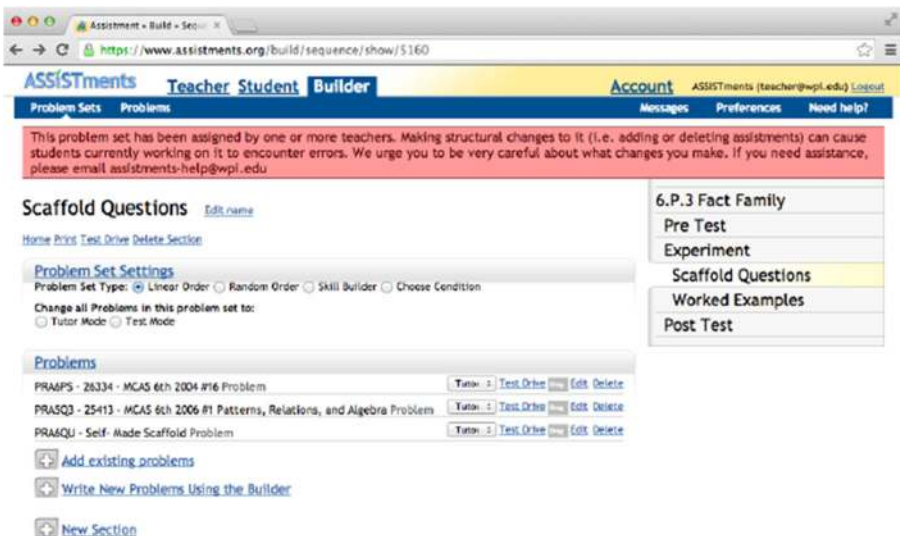**Fig. 5** ASSISTments problem set editor for the section called "Scaffold Questions," which is a problem set for one of the conditions in our example experiment. Since this section is of type "Linear Order," students will see the three ASSISTments PRA6PS, PRA5Q3 and PRA6QU consecutively. The ASSISTments in the "Worked Example" condition had the same questions but their feedback was of the "worked example" type

feedback for the assignment: a worked example or scaffolding questions. One of those problem sets is shown in the problem set editor in Fig. 4. Note the "ChooseCondition" shown in Fig. 4 that allows a researcher to add the randomization needed for a RCT (Fig. 5 shows the "scaffolding questions" condition).

Once the studies are created, we reach out to the teachers asking for volunteers to run these problem sets. We also make the problem sets that we call our ASSISTments Certified Problem Sets. We originally organized the content by the Massachusetts States Frameworks, but we now label them with the Common Core State Standards which helps the growing number of teachers from outside of Massachusetts who use the system. We were not present in classrooms for the running of these studies. By 2009, 186 students had used the problem sets built for the study, which was enough data to find reliable differences between the conditions. While many recent papers have reported reliable gains in learning by the use of worked example, interestingly, we found a reliable decrease in student learning. In the paper we speculate that this finding might be explained by the fact that we used worked examples as a feedback mechanism while others replaced problem solving episodes with worked examples: worked examples just might not make a useful form of feedback. Since then, over 1,000 students have used these problem sets and seen either the worked example or the scaffolding questions.

We would like to provide a second example that explains how our ecosystem allows us to run this type of experiment in a short amount of time. In 2013, I suggested to a first-year graduate student that she conceive a study, design the materials, get students to use the material, analyze the results, and write a paper. She decided to do a study comparing text versus video hints. She took three standard Pythagorean Theorem items already in use in ASSISTments with text hints and videoed herself reading the hints in front of a white board. As part of nurturing our ecosystem, many of our students visit a school once a week to see the system in action and stay connected to the teachers and schools. While in the schools, she had students solve her problems, being randomly exposed to video or text hints. We also asked teachers not in the area to use the problem set. The study had some promising findings, though preliminary, and was worthy of submission (later published as Ostrow and Heffernan 2014a, b). The point of mentioning this study is that the ecosystem, which includes the platform, allowed us to rapidly conduct such a study. But can we do this with other scientists who don't work for WPI?

We have just completed our first such trial. Dr Bethany Rittle Johnson suggested her graduate student, Ms Fyfe, do a study in ASSISTments. She had a phone conversation with Ms Heffernan who explained to make an account, and build her study. Ms Heffernan reached out to our community of teachers and found a teacher who was interested in doing the study. After two more phone calls, Ms Heffernan approved the study, and gave the problem set number to the teacher. Later, Ms Heffernan has the data anonymized and sent to Ms Fyfe. She found the results meaningful, and is preparing a publication. She approached us with the request to do another study, and we asked her why she was eager to continue using ASSISTments. She responded that, "There were two areas in which I saved a lot of time: 1. School negotiations - (Getting schools and teachers on board and getting student consent is a very time-consuming, effortful process. Using ASSISTments students was very quick and efficient.) 2. Data collection - (Many projects I run consist of one-on-one tutoring studies that take at least one semester to complete. Getting data from hundreds of students at once via an ASSISTments assignment was very fast!)"

In addition to Ms Fyfe, there are many studies that have been done and many in the pipeline (Heffernan 2014b). To give a sense of some of the other types of studies we have done, or are currently doing, a complete web site with many examples is posted (Heffernan 2014a). We list a few of the research questions that have done, or are currently undergoing: 1) Does seeing humor during a problem set help learning? 2) Do motivational videos from your teacher help? 3) Does seeing a video of your teacher help encourage completion of summer work? 4) Do hints work better than worked examples? 5) Is it better to see two examples showing the same strategy together or on two different problems? 6) Can we support students who are really struggling by giving them instruction before we continue the practice? and 7) Is there a difference in video that is just a screenshot with someone's voice compared to a video where you can see the person?

This brings up the question of how we can work with other scientists. In this next section we will discuss the idea that ASSISTments is a shared scientific instrument (like an electron microscope that many researchers share to do their own science).

ASSISTments as a Shared Scientific Instrument While we are certainly not the only website that does randomized controlled trials (hundreds of commercial companies run, what it usually called in industry circles, A/B experiments comparing A to B), we have probably taken this idea of a platform designed for experimentation farther than many in the educational technology space. We know of several university-created non-commercial sites like EdX.org, LonCapa.org, and WebWork.org that allow homework to be done online, but none of them have built in support to allow end-users to create RCTs. We now think of ourselves more as the creators, and maintainers, of an instrument that others can use to do their science. The scientists who use the tool do not have to be the ones who made the tool. The scientists who created those instruments don't have to be the ones designing the experiments that use their tool. Instead they solicit studies from other scientist to run on the instrument. They have committees to decide on the studies to be implemented in the limited time available. Luckily for us, ASSISTments can be used to run many experiments at the same time. This realization, that our value might not be in coming up with the studies, but in helping others get their studies run, has encouraged us to collaborate with other scientists from universities such as Columbia, CMU, Harvard, Notre Dame, SRI, Stanford, Texas A&M, UBC, UC-Berkeley, Univ. of Colorado-CS, Univ. of Illinois-Chicago, UMASS, Univ. of Kansas, and Univ. of Maine. Initially those collaborations where with Heffernan as co-author, but we seem to have moved beyond that now.

In anticipation of a NSF grant recommended for funding, we hosted a webinar July 28th, 2014 and had 58 researchers show up to learn how to construct their own studies, and use our subject pool (HHeffernan et al. 2014). Since that webinar, three new studies are now "in the process" by "outsiders" (Heffernan and Heffernan 2014).

Our system is set up so that any researcher can use it for their own purposes. Without technology a typical PhD student in mathematics education might spend a year running a study from conception, through materials creation, to pretest and posttest and then finally write up. We can automate some of that process. Once they create their study on the platform, they can run it with teachers they recruit, or more commonly, we run it on subject pool of current ASSISTments users. WPI's IRB has approved the running of such studies, requiring external researchers to seek their own approval, under the de-

identified exemption (WPI IRB 2014). Meanwhile, WPI's IRB approves of the running the studies that are comparing normal instructional practices under the "comparing normal instructional practices" exemption category. The researcher manual, and IRB forms are available (Ostrow and Heffernan 2014a, b).

Lesson Two: Put the Teachers in Charge

How is it that we get the schools to want to participate in these studies? This bring us to the second lesson learned - put teachers in charge. We determined we needed teachers and students to be using our tool regularly. Therefore we put the teachers in charge allowing them to use the tool on a daily basis. We believe however that teachers won't want to use our system unless it's flexible enough to let them use it to support what is going on in class. We also believe that if the teacher is involved with the workings of the tool they will be able to step in and offer the human help that is often needed with instruction.

Putting Teachers in Charge

Everything the student does is assigned by their teacher. In many ways the list of problem sets in ASSISTments is a replacement for the assignment from the textbook or the traditional worksheet with questions on it. This way the teachers do not have to make a drastic change to their curriculum in order to start using the system. But more importantly they can make more sense of the data they get back since they are the ones who selected and assigned the problems.

This is in contrast to the idea of an artificial intelligence automatically deciding what problem is best for each student. While this is a neat idea, it takes the teacher out of the loop and makes the computer tutorial less relevant to what is going on in the classroom. For instance, the Cognitive Tutor presents this as a key feature. The idea that different students can proceed on their own pace has great appeal to school leaders and makes sense to parents. What is sometimes not appreciated is how that design decision sometimes has unintended consequences. Continuing with Cognitive Tutor as an example, schools are told that it's easy to use as student can automatically move on to the next topic. (Inside ASSISTments, they cannot do so). If you visit one of these computer labs, you will notice that some students are on Chapter 23 while others are still on Chapter 5, while back in the classroom the students are on Chapter 10. This has the potential to create a disconnect in what is being done on the computer and what is being done in the classroom. We have a different design in mind that requires teacher involvement on a daily basis to make online assignments that keep what the students are doing on ASSISTments in line with what is happening in the classroom. For instance, we don't spend much time thinking about computer lab use cases, but instead think of nightly homework as our use case. If some kids on their skill builder homework have to spend more time to demonstrate they learned the skill of the day, we think this is a good way of making something adaptive while keeping the teacher in charge.

To be clear, we are not saying there is one answer to this. There are downsides to having the teacher in charge. It is hard for them to make 100 different decisions one for

each of their students. The most important upsides of the teacher being in charge of sequencing is that the teachers can understand how the system is working, and can better integrate the information with their existing classroom practices.

Putting Teachers in Charge their Own Content

If our content was the only content available then our users would only be teachers who wanted that content. Instead, our builder tool is available to any teacher with an account. We have social studies teachers using the system because they value immediate feedback. Nothing is available to them in ASSISTments certified content; they have to build their questions themselves. Even middle school math teachers who have a treasure trove of content built by the ASSISTments team still want to build their own content.

It is important to note that most of the content built by teachers has only what we call correctness feedback. Teachers find that even without hints they get data, including percent correct and common wrong answers while students get immediate information on whether they are right or wrong on each problem and are given the chance to try again. They are right to think that even this minimal level of assistance is valuable. We have conducted a study (Kelly et al. 2013a) that shows that even this limited feedback leads to large gains in students' learning. Specifically, we wanted to estimate the effects of immediate feedback on homework compared to business-as-usual where students do not get feedback until the next day. Students were randomly assigned to either a control condition or an experimental condition. Students in the control condition were given homework to replicate the traditional scenario of getting feedback the next day, whereas students in the experimental condition received correctness-only feedback as they did their homework. In Kelly et al. (2013b), we found a reliable increase in student knowledge when students got immediate feedback on their homework, compared to next-day feedback. The effect size was half a standard deviation. We also showed a benefit in having the teacher review homework using the data.

There is a feature embedded in the item report that allows teachers to immediately write feedback for common wrong answers and some teachers write their own hints. The biggest benefit of the builder is that teachers can create their own content. To determine what percentage of teachers take advantage of this tool, we analyzed the teachers using ASSISTments on a randomly selected school day and found that 76 % of these teachers had, at some point, created their own content and 29 % had created content that day. We have found that many teachers want to create their own content. It might sound obvious, but the key seems to be giving them simple but powerful tools that allow them to do that. This then allows us to implement studies without disrupting the classroom routine.

Putting Teachers in Charge of Students' Data

The reports we have built in conjunction with our teachers are a key way of keeping them in charge and getting them to see the value of adopting ASSISTments. Skill builders have their own report that allows teachers to know if a student mastered the

skill, how many problems it took, and how much time. For normal problem sets, teachers can see the percent correct for each student and each item. We think (and our experience working with teachers verifies) that since the teacher was the one who assigned the problems the report makes perfect sense to them. They find the discussions over the most difficult problems and common wrong answers so compelling they keep using ASSISTments to assign their work.

In the study on feedback and homework, Kelly analyzed the way that she, as a teacher, used the data from ASSISTments to go over the homework. The study reported that students learned a great deal from teachers reviewing homework with the data from the reports, Fig. 6 and 7 (Kelly et al. 2013b). For example, the teacher used the report in Fig. 7 to see that on the 4th question, only 27 % of her students answered the question correctly and 56 % of students provided the wrong answer of 1/9^10. This suggested that the students have a similar misconception which she can then address. The control and experiment video conditions are available at (http://www.aboutus.assistments.org/learning-increase-with-homework.php). The Experimental Condition video reveals that the teacher could use that information to prompt a meaningful discussion. The growth from posttest 1 (immediately following the homework, before the in-class review) to posttest 2 (after the in-class review) showed reliable gains in student learning, due to the classroom discussion informed by the ASSISTments report, from 68 % correct to 81 % correct (Kelly et al. 2013a).

Teachers are Put in Charge of the Adaptively of ASSISTments

When we say teachers should be put in charge, in our opinion this implies that they can be in charge of how the system should differentiate between students. This is best illustrated with our Data Driven Rule feature. Once a teacher has assigned a problem set she can use the Data Driven Rule to respond to the data. She can assign any problem set to all the students who got a problem wrong or to all the students who got a problem

Fig. 6 This is the item report from K. Kelly's study (Kelly et al. 2013a&b)

right. For example, teachers give extra practice on the skill to students who got the problem wrong (we have individual problem sets for over 140 skills in math).

Another use case is to assign more challenging problems to the students who got the problem right while the teacher then works in small groups with the students who got the problem wrong. These data driven actions are saved by ASSISTments and the teacher can use them the next time the assignment is used (see Fig. 7).

This illustrates how teachers are kept in charge - they choose how the adaptation should work - but the computer delivers the assignment to the students. This is especially useful when a teacher has just a few students in need of additional help or challenge. We could have made the data driven rules function work automatically, but that would have violated our philosophy of putting the teacher in charge. We think they might not have paid as much attention to students work if that were the case. If the teacher assigned the problems, the teacher is more likely to look at the data.

While we intended teachers to use the report, for the writing of this paper we looked to see how often teachers looked at their reports. During the 2013–2014 school year teachers made a total of 13,442 problem sets, and there was 10,017 item report unique views. Meaning that most of the time (74.5 % of the time) a teacher pulled up the "Item Report" (as shown in Fig. 7) that shows details on student work. This number underestimates teacher viewing, as for teachers are emailed information on which students did their work, and what were the hard questions, so they might not feel the need to see common wrong answers or the details of which students got which questions wrong. We are continually working to increase the use of Item Reports. Our goal is to have all teachers not only looking the brief summary in the email, but also viewing and utilizing the complete reports.

Student/Problem [Unanonymize]	Average Data driven	#213205 Data driven	#213206 Data driven	#213207 Data driven	#213208 Data driven
Problem Average	60%	90%	100%	100%	27%
Common Wrong Answers					1/9^10, 56%
Correct Answer(s)		1/5^3	1/7^8	1/6^12	1/3^10
XXXXX :	70%	✓ 1/5^3	✓ 1/7^8	✓ 1/6^12	✓ 1/3^10
XXXXX :	65%	✓ 1/5^3	✓ 1/7^8	✓ 1/6^12	✗ 1/9^10
XXXXX :	75%	✓ 1/5^3	✓ 1/7^8	✓ 1/6^12	✗ 1/3^6
XXXXX :	50%	✓ 1/5^3	✓ 1/7^8	✓ 1/6^12	✗ 6561

Fig. 7 The teacher report showing the data driven button that can be used toretain information for the future

We think this is important as we believe that while computers can help, there will always be times when the student is so confused he is better off getting the help of a human. We think the ITS field tends to overestimate how effective the computer can be and not value the role of the teachers as much as we think it should be valued.

Summary of Putting Teachers in Charge

We try to put teachers in charge, but we want to acknowledge there are tradeoffs. We acknowledge that our design philosophy has advantages and disadvantages. On the positive side, we think we have increased teachers' adoption because of the fact that ASSISTments is flexible and teachers can do what they want with it. On the downside, due to the flexibility there is more to learn for a teacher, and more work on a daily basis in "tending" to the computer to make assignments or writing content. Additionally, the cost in terms of extra professional development is probably quite substantial. While there is no data we can point to, it seems reasonable to assume that ASSISTments requires more time for the teacher to learn. Ken Koedinger (Koedinger who helped create both Cognitive Tutor and ASSISTments) described cognitive tutor as "a more complete turn-key solution" while ASSISTments is "more flexible". That flexibly requires/allows (depending upon your interpretation) more customization (Koedinger and Martineua 2010).

This may be why only a tiny fraction of the 500 teachers who created accounts after the New York Times (Paul, 2012) did a story on ASSISTments turned into consistent users. We aren't sure what to compare this to but it's certainly the case that we want to lower the learning curve for teachers. Cristina (who is in charge of teacher training) has noticed that only a few teachers begin to use ASSISTments because they see the big picture of using data to inform their instruction; most just see it as a tool ease the burden of grading. She has found however that once a teacher gets past the learning curve for the system they start to think formatively about their data, they bring the data to department meetings and use it for curriculum development. This side benefit of improving teaching as well as learning is exciting. Even if an experienced ASSISTments user has to stop using the tool, we hope they will be a better teacher because they now think of data as a tool for teaching.

While we have argued that the success of ASSISTments is, in part, due to the "teacher in charge" design philosophy, we do feel there is a lot more that can be done to improve the system and make it easier for teachers to use and share. We have even toyed with the idea of building a much more prescriptive version of ASSISTments that would tell teachers exactly what to do each day, but we have resisted such temptations. I am sure some districts will use the platform and their professional development time to train teachers in a more prescriptive way of teaching, while using ASSISTments, and that is fine with us. I guess we think of ASSISTments more as a technology platform/service and less as a complete product with a set ways for teachers to implement it.

Our goal is to figure out how to use computers, and all they have to offer, to maximize learning. We believe the best way to do this is to combine the best of what teachers do well and the best of what computers do well. Computers are best at patience, bookkeeping and being there for every student at the moment they need feedback. Does this mean that we should replace teachers with computers? Absolutely not. Every computer system will fail in certain ways and no system can solve all the problems all the time. We think, teachers are best for motivating students, and

providing conceptual understanding to help students who are confused. Our goal is to help good teachers become better, not to serve as a surrogate math instructor to compensate for bad teachers.

Concluding Thoughts

The ASSISTments "sweet spot" is simple. Teachers can write their own questions and their students get feedback. We can call this a simple quizzing system. We have also added features that teachers can use separate from other features, depending upon their interest. If they want to assign skill builders they can. If they want to use ARRS they can. If they want to use EdRank they can turn that on. The teacher is in charge, not the computer. By building all this we can get studies run that help the learning sciences community.

So where do we think we see our project going?

Our Future Plans - Wikipedia for Questions and Feedback

We would like to step back and reflect upon where we are trying to take ASSISTments. We currently have about 25,000 vetted questions created by WPI, CMU and our small number of university partners. But over 100,000 questions have been written by teachers. We are hoping to create something like Wikipedia, not for encyclopedia entries, but for educational questions and their feedback. We hope to have every commonly used math textbook supported, so students can get immediate feedback. In Maine we have 30 of the most commonly used 7th grade math textbooks already being used; the students still need their book as we don't violate the textbook publisher copyright. However, when teachers write hints they are not owned by the publisher. We hope that as teachers curate textbooks, they will add hints and feedback for common misconceptions. We are working on a design to help crowd-source those hint and feedback messages from students, so they can help each other, but in ways that constrain them so that the teacher can see who could get it correct without assistance.

Speaking of sharing content, our most successful teacher, is at the prestigious Boston Latin High School, Eric Simoneau, who won his own NSF grant to build online materials to support AP Statistics, using ASSISTments. He has written over 500 problems for AP Statistics. (at Stats4STEM.org) He has shared them with 400 other AP statistics teachers, (so each year our server gets slammed the weekend before the AP test and tens of thousands of questions are solved). We want to help teachers like this share their work, for free, with the rest of the world. This is why ASSISTments is currently, and will forever be, free.

When we mention Wikipedia that does not mean we want anyone to edit anything at anytime. We think student vandalizing questions is a potential problem, as well as getting "answer keys," so we created ASSISTments so that no one can change a teacher's questions, except that teacher, but anyone can make a copy of a problem and then assign that to their students. This allows for collaboration between teachers. Of course comparing ourselves to Wikipedia, one of the most useful and successful sites on the planet is a touch of hubris and a vision we are not likely to achieve, but that is our goal.

But in addition to providing a valuable public service, we also want to leverage the ASSISTments platform as a scientific platform. We want hundreds of experiments running inside of ASSISTments so we can learn as much as possible about learning.

We know that the millions of students each night who are not getting feedback on their homework are in need of better, smarter homework. We think the world will look back at what schools are doing now and think of it as "educational malpractice." We know we can do better.

Finally, we encourage others to use our tool to do their own science; we don't want to be authors on your papers. There are too many ideas for us to think we need to do them all!

References

Aleven, V. (2010). Rule-based cognitive modeling for intelligent tutoring systems. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Studies in computational intelligence* (Advances in intelligent tutoring systems, Vol. 308, pp. 33–62). Berlin: Springer. doi:10.1007/978-3-642-14363-2_3.

Aleven, V., McLaren, B. M., & Sewall, J. (2009a). Scaling up programming by demonstration for intelligent tutoring systems development: an open-access website for middle-school mathematics learning. *IEEE Transactions on Learning Technologies, 2*(2), 64–78.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2006). The cognitive tutor authoring tools (CTAT): preliminary evaluation of efficiency gains. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Proceedings of the 8th international conference on intelligent tutoring systems.* Taiwan: Jhongli.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009b). A new paradigm for intelligent tutoring systems: example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105–154.

Anderson, J. R., & Pelletier, R. (1991). A developmental system for model-tracing tutors. In B. Lawrence (Ed.), *The international conference on the learning sciences* (pp. 1–8). Charlottesville: Association for the Advancement of Computing in Education.

Anderson, J. R. (1993). *Rules of the mind.* Mahwah: Lawrence Erlbaum.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207.

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human Computer Studies, 68*(4), 223–24.

Blessing, S., Gilbert, S. B., Ritter, S. (2006). Developing an authoring system for cognitive models within commercial-quality ITSs. Proceedings of the 19th Florida Artificial Intelligence Research Society Conference (FLAIRS-19). Press. pp. 497–502.

Bloom, B. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Broderick, Z., O'Connor, C., Mulcahy, C., Heffernan, N., & Heffernan, C. (2012). Increasing parent engagement in student learning using an intelligent tutoring system. *Journal of Interactive Learning Research, 22*(4), 523–550.

Brusilovsky, P., Schwarz, E., & Weber, G. (1996). ELM-ART: an intelligent tutoring system on World Wide Web. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), *Intelligent tutoring systems (Lecture Notes in Computer Science)* (Vol. 1086, pp. 261–269). Berlin: Springer Verlag.

Chi, M., VanLehn, K., Litman, D., & Jordan, P. (2011). An evaluation of pedagogical tutorial tactics for a natural language tutoring system: a reinforcement learning approach. *International Journal of Artificial Intelligence in Education, 21*(1–2), 83–113. doi:10.3233/JAI-2011-014.

Chen, W., Aist, G., & Mostow, J. (2011). Recognizing young readers' spoken questions. *International Journal of Artificial Intelligence and Education*, 21(4).

Corbett, A. (2001). Cognitive computer tutors: solving the two-sigma problem. In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva (Eds.), *Proceedings of the 2001 international conference on user modeling* (pp. 137–147). Berlin: Springer.

D'Mello, S. K., Lehman, B., & Person, N. (2010). Expert tutors feedback is immediate, direct, and discriminating. Proceedings of the 23rd Florida Artificial Intelligence Research Society Conference (FLAIRS-23). AAAI Press. pp. 595–604.

du Boulay, B., Avramides, K., Luckin, R., Martínez-Mirón, E., Méndez, G. R. & Carr, A. (2010). Towards systems that care: a conceptual framework based on motivation, metacognition and affect. *International Journal of Artificial Intelligence in Education* 20(3).

Feng, M. & Heffernan, N. (2010). Can We get better assessment from a tutoring system compared to traditional paper testing? Can we have our Cake (Better Assessment) and eat it too (Student Learning During the Test) In Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining, Pittsburgh, PA. (pp 41–50).

Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). Addressing the assessment challenge in an intelligent tutoring system that tutors as it assesses. *The Journal of User Modeling and User Adapted Interaction., 19*, 243–266.

Feng, M., Roschelle, J., Heffernan, N., Fairman, J. & Murphy, R. (2014) Implementation of an intelligent tutoring system for online homework support in an efficacy trial. In Stefan Trausan-Matu, et al. (Eds) The Proceeding of the International Conference on Intelligent Tutoring 2014. LNCS 8474. pp 561–566.

Gong, Y., Beck, J. & Heffernan, N. (2012). WEBsistments: enabling an intelligent tutoring system to excel at explaining why other than showing how. Proceedings of the 11th International Conference on Intelligent Tutoring Systems, Greece.

Heffernan, N. T (2001). Intelligent tutoring systems have forgotten the tutor: adding a cognitive model of human tutors dissertation. Computer Science Department, School of Computer Science, Carnegie Mellon University.

Heffernan. N. (2014a). ASSISTments: as a researcher's tool: studies retrieved July 28th 2014 from https://sites.google.com/site/assistmentsstudies/home

Heffernan, N. (2014b). Adding research accounts to the ASSISTments platform:randomized controlled studies with thousands of students. An National Science Foundation grant abstract. Retrieved July 28th, 2014 from http://goo.gl/TmlxQv

Heffernan, N. T. (2003). Web-based evaluations showing both cognitive and motivational benefits of the Ms. Lindquist tutor In F. Verdejo and U. Hoppe (Eds) 11th International conference artificial intelligence in education. Sydney, Australia. pp.115-122.

Heffernan, N. T. & Croteau, E. (2004). Web-based evaluations showing differential learning for tutorial strategies employed by the Ms. Lindquist Tutor. In James C. Lester, Rosa Maria Vicari, Fábio Paraguaçu (Eds.) Proceedings of 7th annual intelligent Tutoring Systems Conference, Maceio, Brazil. (pp. 491–500).

Heffernan, N. & Heffernan, C. (2014) Application process to by outsiders submit a study. Retrieved August 7, 2014 from https://sites.google.com/site/assistmentstestbed/creating-a-study/getting-started

Heffernan, N., Heffernan, C., Dietz, K., Soffer, D., Pellegrino, J. W., Goldman, S. R. & Dailey, M. (2012). Improving mathematical learning outcomes through automatic reassessment and relearning. Proceedings of AERA 2012. Vancouver, BC, Canada.

Heffernan, N., Heffernan, C., D'Mello, S., Williams, J. J. and Stoelinga, T. (2014) A webinar explaining how WPI offers to be the midwife to others researchers' ideas. Retrieved August 9, 2014 from https://sites.google.com/site/neilheffernanscv/webinar

Heffernan, N. T., & Koedinger, K. R. (2000). Intelligent tutoring systems are missing the tutor: Building a more strategic dialog-based tutor. In C.P. Rose & R. Freedman (Eds.) Proceedings of the AAAI fall symposium on building dialogue systems for tutorial applications. Menlo Park, CA. pp. 14–19.

Heffernan, N., Militello, M., Heffernan, C., & Decoteau, M. (2012b). Effective and meaningful use of educational technology: three cases from the classroom. In C. Dede & J. Richards (Eds.), *Digital teaching platforms* (pp. 88–102). Columbia: Teachers College Press.

Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education, 20*, 175–195.

Kehrer, P., Kelly, K. & Heffernan, N. (2013). Does immediate feedback while doing homework improve learning. In Boonthum-Denecke, Youngblood(Eds) Proceedings of the twenty-sixth international Florida artificial intelligence research society conference, FLAIRS 2013, St. Pete Beach, Florida. May 22–24, 2013. AAAI Press 2013. p 542–545.

Kelly, K., Heffernan, N., D'Mello, S., Namias, J., & Strain, A. (2013a). Adding teacher-created motivational video to an ITS. Florida artificial intelligence research society (FLAIRS 2013). pp. 503–508.

Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., Goldstein, D. S. (2013b). Estimating the Effect of Web-Based Homework. In Lane, Yacef, Motow & Pavlik (Eds) The Artificial Intelligence in Education Conference. Springer-Verlag. pp. 824–827. (Archived material is at http://web.cs.wpi.edu/~nth/PublicScienceArchive/Kelly.htm). A longer version is available as Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., Goldstein, D. S. (2013). Estimating the effect of web-based homework. in formative feedback in interactive learning environments (FFILE) Workshop. Held at the artificial intelligence in education conference, Memphis.

Kim, R, Weitz, R., Heffernan, N. & Krach, N. (2009). Tutored problem solving vs. "Pure": worked examples In N. A. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st annual conference of the cognitive science society (pp. 3121–3126). Austin, TX: cognitive science society.

Koedinger, K. R., Aleven, V., Heffernan. T., McLaren, B. & Hockenberry, M. (2004). Opening the door to non-programmers: authoring intelligent tutor behavior by demonstration. In James C. Lester, Rosa Maria Vicari, Fábio Paraguaçu (Eds.) Proceedings of 7th annual intelligent tutoring systems conference, Maceio, Brazil. (pp.162-173).

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30–43.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*(6161), 935–937. doi:10.1126/science.1238056.

Koedinger, K., & Martineua, J. (2010) Technology capabilities to assess student cognition and learning in next generation classroom formative assessments. Retrieved from http://www.ctb-e.com/emails/pdfs/Asst_Webinar_11_30_10_final.pdf Slide 40.

Koedinger, K., McLaughlin, E., & Heffernan, N. (2010b). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research Baywood Publishing, 4*, 489–510.

Lebeau, J., Fortin, M., Paquette, L., Mayers, A. (2009) From cognitive to pedagogical knowledge models in problem-solving ITS frameworks. In Dimitrova, Mizoguchi, du Boulay and Graesser (Eds.) Proceedings of the conference on artificial intelligence in education. pp. 731–733

Lester, J., Rowe, J., & Mott, B. (2013). Narrative-centered learning environments: a story-centric approach to educational games. In C. Mouza & N. Lavigne (Eds.), *Emerging technologies for the classroom: a learning sciences perspective* (pp. 223–238). Manhattan: Springer US.

Li, S., Xiong, X., Beck, J. (2013). Modeling student retention in an environment with delayed testing. The sixth international conference on educational data mining. Memphis, TN. pp. 328–329

Miller, G., Zheng, Y., Means, B., & Van Brunt Miller, B. (2013). Next generation learning challenges wave II, Retrieved on August 1st 2014 from http://goo.gl/V6pGUo.

National Science Foundation (NSF, 2008). Partnership implementing math and science education: assisting middle school use of tutoring technology. A Grant written by Heffernan funded by the NSF. Abstract retrieved on Jan 2013 from http://www.nsf.gov/awardsearch/showAward?AWD_ID = 0742503.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. Cognition and instruction. 7 pp. 197–244.

Mendicino, M., Razzaq, L., & Heffernan, N. T. (2009). Comparison of traditional homework with computer supported homework. *Journal of Research on Technology in Education, 41*(3), 331–359.

Mitrovic, A., & Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. *International Journal of Artificial Intelligence in Education, 10*(3), 238–256.

Murray, T. (1999). Authoring intelligent tutoring systems: an analysis of the state of the art. *International Journal of Artificial Intelligence in Education, 10*, 98–129.

Ohlsson, S. (1992). Constraint-based student modeling. *Journal of Artificial Intelligence in Education, 3*(4), 429–447.

Ostrow, K., & Heffernan, N. T. (2014) Testing the multimedia principle in the real world: a comparison of video vs. Text feedback in authentic middle school math assignments. In John Stamper et al. (eds) Proceedings of the 7th international conference on educational data mining. pp. 296–299.

Ostrow, K., & Heffernan, C. (2014) How to create controlled experiments in ASSISTments. A manual for researchers. Retrieved on August 9, 2014 from https://sites.google.com/site/assistmentstestbed/

Pardos, Z., Dailey, M., & Heffernan, N. (2011). Learning what works in ITS from non-traditional randomized controlled trial data. *International Journal of Artificial Intelligence in Education, 21*, 47–63.

Pardos, Z. & Heffernan, N. (2012). Tutor modeling vs. Student modeling. Proceedings of the twenty-fifth international Florida artificial intelligence research society conference. Invited talk. Florida Artificial Intelligence Research Society (FLAIRS 2012). pp 420–425.

Paul, A. M. (2012). The Machines are Taking over. New York Times, Education Section. Retrieved on Sept 12, 2014 from http://www.nytimes.com/2012/09/16/magazine/how-computerized-tutors-are-learning-to-teachhumans.html?pagewanted=all&_moc.semityn.www&_r=1&

Person, N. K, Graesser, A. C., Kreuz, R. J., Pomeroy, V. & Tutoring Research Group. (2001). simulating human tutor dialog moves in autotutor. *International Journal of Artificial Intelligence in Education*, 12.

Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N.T., Koedinger, K. R., Junker, B., Ritter, S., Knight, A., Aniszczyk, C., Choksey, S., Livak, T., Mercado, E.,Turner, T.E., Upalekar. R, Walonoski, J.A., Macasek. M.A. & Rasmussen, K.P. (2005). The assistment project: blending assessment and assisting. In C.K.Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.) Proceedings of the 12th artificial intelligence in education, Amsterdam: ISO Press. pp. 555–562.

Razzaq, L. & Heffernan, N.T. (2006). Scaffolding vs. hints in the assistment system. In Ikeda, Ashley & Chan (Eds.). Proceedings of the eighth international conference on intelligent tutoring systems. Springer-Verlag: Berlin. pp. 635–644.

Razzaq, L., & Heffernan, N.T. (2009). To tutor or not to tutor: that is the question. In Dimitrova, Mizoguchi, du Boulay and Graesser (Eds.) Proceedings of the Conference on Artificial Intelligence in Education. pp. 457–464.

Razzaq, L. & Heffernan, N. (2010). Hints: is it better to give or wait to be asked? In Aleven, V., Kay, J & Mostow, J. (Eds) Proceedings of the 10th international conference on intelligent tutoring systems (ITS2010) Part 1. Springer. Pages 349–358.

Razzaq, L., Heffernan, N. T., Lindeman, R. W. (2007). What level of tutor interaction is best? In Luckin & Koedinger (Eds.) Proceedings of the 13th conference on artificial intelligence in education. IOS Press. pp 222–229.

Razzaq, L., Mendicino, M., & Heffernan, N. (2008). Comparing classroom problem-solving with no feedback to web-based homework assistance. In A. Woolf & L. Nkambou (Eds.), *Proceeding of the 9th international conference on intelligent tutoring systems* (pp. 426–437). Berlin: Springer.

Razzaq, L., Patvarczki, J., Almeida, S. F., Vartak, M., Feng, M., Heffernan, N. T., & Koedinger, K. (2009b). The ASSISTment builder: supporting the life-CYCLE of ITS content creation. *IEEE Transactions on Learning Technologies Special Issue on Real World Applications of Intelligent Tutoring Systems, 2*(2), 157–166.

San Pedro, M., Baker, R., Gowda, S., & Heffernan, N. (2013). Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. In Lane, Yacef, Motow & Pavlik (Eds) The artificial intelligence in education conference. Springer-Verlag. pp. 41–50.

Sao Pedro, M., Gobert, J., Heffernan, N. & Beck, J. (2009). In N.A. Taathen & H. van Rjin (Eds.) Comparing pedagogical approaches for teaching the control of variables strategy. Proceedings of the 31st annual conference of the cognitive science society Austin, TX.

Shrestha, P., Wei, X., Maharjan, A., Razzaq, L., Heffernan, N.T., & Heffernan, C., (2009). Are worked examples an effective feedback mechanism during problem solving? In N. A. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st annual conference of the cognitive science society (pp. 1294–1299). Austin, TX: *Cognitive Science Society.*

Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N., Razzaq, L. Dailey, M. O'Connor, C.& Mulchay, C. (2011). Feedback during web-based homework: the role of hints In Biswas et al. (Eds) Proceedings of the artificial intelligence in education conference 2011. pp. 328–336.

Soffer, D., Das, V., Pellegrino, G., Goldman, S., Heffernan, N., Heffernan, C.,& Dietz, K. (2014) Improving long-term retention of mathematical knowledge through automatic reassessment and relearning. American educational research association (AERA 2014) Conference. Division C - learning and instruction/section 1c: mathematics. PDF, poster nominated for the best poster of the session.

Suraweera, S. & Mitrovic, A. (2004). An intelligent tutoring system for entity relationship modelling. *International Journal of Artificial Intelligence in Education*, 14(3).

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J.A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., and Wintersgill, M. (2005). The andes physics tutoring system: lessons learned. *International Journal of Artificial Intelligence and Education, 15* (3).

VanLehn, K., Siler, S., & Murray, C. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249.

Walonoski, J., & Heffernan, N. (2006). Prevention of off-task gaming behavior in intelligent tutoring systems. In A. Ikeda & Chan (Eds.), *Proceedings of the 8th international conference on intelligent tutoring systems* (pp. 722–724). Berlin: Springer.

Wang, H. C., Rose, C. P., & Chang, C. Y. (2011). Agent-based dynamic support for learning from collaborative brainstorming in scientific inquiry. *International Journal of Computer Supported Collaborative Learning, 6*(3), 371–396.

Whorton, S. (2013). Can a computer adaptive assessment system determine, better than traditional methods, whether students know mathematics skills? Masters thesis, computer science department. Retrieved on Jan 10, 2014 from http://www.wpi.edu/Pubs/ETD/Available/etd-041913-095912/

Woolf, B. P. (2009) *Building Intelligent Tutoring Systems*. Morgan Kaufman.

Woolf, B.P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D., Dolan, R., & Christopherson, R.M. (2010). The effect of motivational learning companions on low-achieving students and students with learning disabilities. Proceedings of the international conference on intelligent tutoring systems, Pittsburgh, PA.

WPI IRB (2014) Terms of use of using the platform to to run a study. Worcester polytechnic institute's institutional review board. Retrieved Augusts 7th, 2014 from https://sites.google.com/site/assistmentstestbed/terms-of-use

Xiong, X. & B0eck, J. (2014) A study of exploring different schedules of spacing and retrieval interval on mathematics skills in ITS environment. In Stefan Trausan-Matu, et al. (Eds) The Proceedings of the International Conference on Intelligent Tutoring 2014. LNCS 8474. Pages 504–509.