# Student modeling in an Intelligent Tutoring System

Mingyu Feng[1], Neil Heffernan[1], Kenneth Koedinger[2]

{mfeng|nth}@wpi.edu, Koedinger@cmu.edu

[1]Department of Computer Science, Worcester Polytechnic Institute, USA

[2]Human Computer Interaction Institute, Carnegie Mellon University, USA

**Abstract**

Intelligent tutors have often been used mainly to teach students. In the ASSISTments project, we have emphasized using the intelligent tutoring system as an assessment system that just so happens to provide instructional assistance during the test. In this chapter we review and summarize some of the main studies we have done with the system. Usually its believed that assessment get harder if students are allowed to learn during the test, as its then like try to hit a moving target. So our results are surprising, that by providing tutoring to students while they are assessed we actually improve the assessment of students' knowledge. We also review out attempts to give teachers feedback based on fine grained skill models. Overall, we conclude that using intelligent tutoring systems to do assessment seems like a reasonable way of dealing with the dilemma that every minute spent testing students takes time away from instruction.

**Key word**: Intelligent Tutoring System, online tutoring, assessment, student modeling

## Introduction

In many States there are concerns about poor student performance on new high-stakes standards based tests that are required by No Child Left Behind Act (NCLB). Partly in response to this pressure, and partly because teachers, parents, and other stakeholders want and need more immediate feedback about how students are doing, there has recently been intense interest in using periodic benchmark tests to predict student performance on end-of-year accountability assessments (Olson, 2005). Some teachers make extensive use of practice tests and released items to target specific student knowledge needs and identify learning opportunities for individual students and the class as a whole so that their instruction could be data-driven. However, such formative assessments not only require great effort and dedication, but they also take valuable time away from instruction. Thereby, limited classroom time available in middle school classes compels teachers to choose between time spent assisting students' development and time spent assessing students' abilities. Critics of NCLB are calling the bill "No Child Left Untested" emphasizing the negative side of assessment, in that every hour spent assessing students is an hour lost from instruction. Or does it have to be? What if we better integrated assessment into the classroom, and we allowed students to learn during the test? Noticing the dilemma, Heffernan at Worcester Polytechnic Institute and his colleagues at Carnegie Mellon University started to build a system ("ASSISTments"[i]) to help resolve this dilemma. Traditionally the two areas of testing (i.e. Psychometrics) and instruction (i.e., math educational research and instructional technology research) have been separate fields of research with their own goals. Therefore, in order to put them together the solution here must involve a way whereby students can take an assessment, but at the same time, make sure that students' time is spent primarily on learning. It should be able to provide an accurate prediction of student performance on the standardized tests so that teachers have an idea of how their students will perform on end-of-year assessment. Meanwhile, it will present a more fine-grained evaluation of student abilities so the teachers will be able to use this detailed feedback to tailor their instruction to focus on the particular difficulties identified by the system.

In this chapter, we will describe how the solution has been achieved by providing both assistance and assessment in an integrated fashion. In the first section, we focus on giving an overview of the ASSISTments system, including the structure of an ASSISTment, the problem sequencing, the teacher reports, the authoring tools, content development and usage and also the evidence showing the effectiveness of tutoring in ASSISTments. The second section is devoted to student modeling and performance estimation in the ASSISTments system. We will describe how we improve the accuracy of assessment by tracking how much assistance students need. Furthermore, our fine-grained cognitive models that map each question to a few knowledge components allow us to more accurately predict these scores. We conclude the chapter in the third section.

## Background on the ASSISTments System

The ASSISTments project is joint research conducted by Worcester Polytechnic Institute and Carnegie Mellon University and is funded by grants from the Department of Education, the National Science Foundation, and the Office of Naval Research. The ASSISTment project's goal is to provide cognitive-based assessment of students while providing tutoring content to students.

The ASSISTment system aims to assist students in learning the different skills needed for the Massachusetts Comprehensive Assessment System (MCAS) test or (other state tests) while at the same time assessing student knowledge to provide teachers with fine-grained assessment of their students' knowledge; it assists while it assesses. The system assists students in learning different skills through the use of scaffolding questions, hints, and incorrect messages (or buggy messages) (Razzaq et al., 2005). Assessment of student performance is provided to teachers through real-time reports based on statistical analysis. Using the web-based ASSISTment system is free and only requires registration on our website; no software need be installed. Our system is primarily used by middle- and high-school teachers throughout Massachusetts who are preparing students for the MCAS tests. Currently, we have over 3000 students and 50 teachers using our system as part of their regular math classes. We have had over 30 teachers use the system to create content.

Though ASSISTments is a derivative of Cognitive Tutor (Anderson et al., 1995), it is built for different anticipated classroom use from the Cognitive Tutor. Cognitive Tutor students are intended to use the tutor two class periods a week. Students are expected to proceed at their own rate letting the mastery learning algorithm advance them through the curriculum. Some students will make steady progress while others will be stuck on early units. There is value in this in that it allows students to proceed at their own paces. One downside from the teachers' perspective could be that they might want to have their class all do the same material on the same day so they can assess their students. ASSISTments were created with this classroom use in mind. ASSISTments were created with the idea that teachers would use it once every two weeks as part of their normal classroom instruction, meant more as a formative assessment system and less as the primary means of assessing students.  Cognitive Tutor advances students only after they have mastered all of the skills in a unit. We know that some teachers use some features to automatically advance students to later lessons because they might want to make sure all the students get some practice on quadratics, for instance.

We think that no one system is “the answer” but that they have different strengths and weaknesses.  If the student uses the computer less often there comes a point where mastery learning based program (i.e., the Cognitive Tutor) may be behind on what a student knows, and seem to move along too slowly to teachers and students.  On the other hand, a weakness of ASSISTments is that does not offer mastery learning and adaptive activity selection, so if students struggle, it does not automatically adjust. It is assumed that the teacher (and not the computer system) will decide if a student needs to go back and look at a topic again.

## The Structure of an ASSISTment

Koedinger et al. (2004) introduced pseudo-tutors which mimic cognitive tutors but are limited to a single problem. The ASSISTment system uses a further simplified pseudo-tutor, called an ASSISTment, where only a linear progression through a problem is supported which makes content creation easier and more accessible to a general audience.

An ASSISTment consists of a single main question (a.k.a. original question) and a tutoring session for assistance. As students working in the system, the main question will be presented first and can be treated as an assessment task for which students need to submit an answer. In contrast to a traditional testing environment, students can request assistance if they don’t know how to answer the question, though it is generally thought to be pedagogically more desirable that a student submits a thoughtful answer before accessing the tutoring. For any given problem, assistance to students is available either in the form of a hint sequence or a set of scaffolding questions. Hints are messages that provide insights and suggestions for solving a specific problem, and each hint sequence ends with a *bottom-out* hint which gives the student the answer. Scaffolding questions are designed to lead the student one-step-at-a-time to the solution and each step addresses specific skills needed to answer the original question. Students must answer each scaffolding question in order to proceed to the next scaffolding question. When students finish all of the scaffolding questions, they may be presented with the original question again to finish the problem. Each scaffolding question also has a hint sequence to help the students answer the question if they need extra help. Additionally, constructive feedback called *buggy* messages are provided to students if certain anticipated incorrect answers are selected or entered, otherwise a generic feedback message will be shown. For problems without scaffolding, a student will remain in a problem until the problem is answered correctly and can ask for hints which are presented one at a time. If scaffolding is available, the student will be programmatically advanced to the scaffolding questions in the event of an incorrect answer. The flowchart in Figure 1 portrays the interaction between an ASSISTment and a student.

[INSERT FIGURE 1 ABOUT HERE: A flowchart that portrays interaction between an ASSISTment and a student]

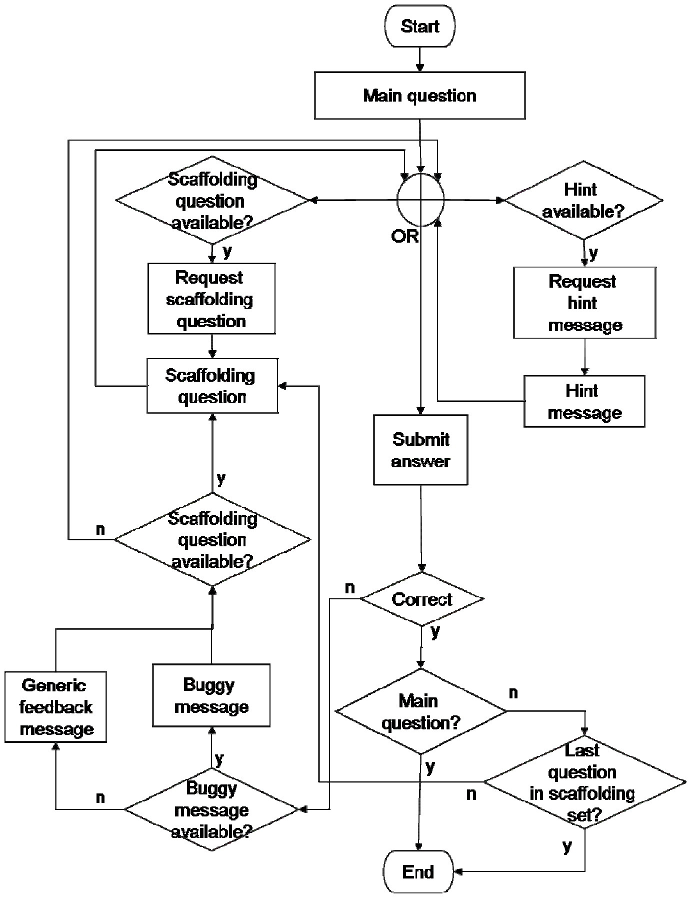



**Figure 1. A flowchart that portrays interaction between an ASSISTment and a student**

The ASSISTments assume that students may know certain skills and rather than slowing them down by going through all of the scaffolding first, ASSISTments allow students to try to answer questions without showing every step. This differs from Cognitive Tutors (Anderson et al., 1995) and Andes (VanLehn et al., 2005) which both ask the students to fill in many different steps in a typical problem. We prefer our scaffolding pattern as it means that students get through items that they know faster and spend more time on items they need help on. It is not unusual for a single Cognitive Tutor Algebra Word problem to take ten minutes to solve, while filling in a table of possibly dozens of sub-steps, including defining a variable, writing an equation, filling in known values, etc. We are sure, in circumstances where the student does not know these skills, that this is very useful. However, if the student knows most of the steps this may not be pedagogically useful.

## 1.2 The Authoring tools

Hints, scaffolds, and buggy messages together help create ASSISTments that are structurally simple but can address complex student behavior and provide appropriate intervention. The structure and the supporting interface used to build ASSISTments (*the authoring tools* or sometimes referred to as *the builder*), shown in **Error! Reference source not found.**, uses common web technologies such as HTML and JavaScript, allowing it to be used on most modern browsers. The authoring tools are simple enough so that users with little or no computer programming experience or cognitive psychology background can use it easily. **Error! Reference source not found.** shows an ASSISTment being built on the left and what the student sees is shown on the right. Content authors can easily enter question text, hints and buggy messages by clicking on the appropriate field and typing; formatting tools are also provided for easily bolding, italicizing, etc. Images and animations can also be uploaded in any of these fields. The builder also enables scaffolding within scaffold questions, although this feature has not been often been used in our existing content.

[INSERT FIGURE 2 ABOUT HERE:  An item in the builder and on the corresponding student screen]

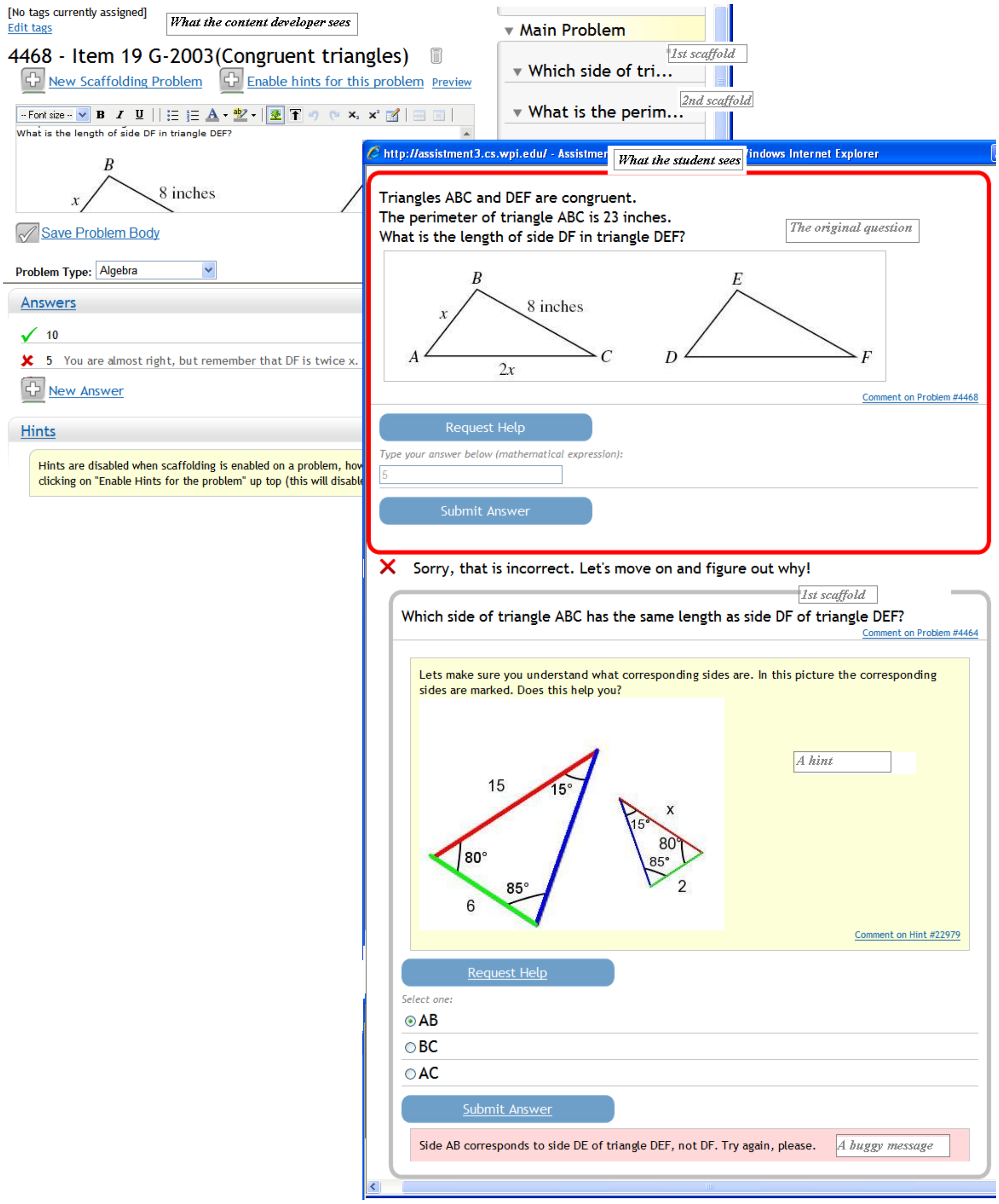



**Figure 2. An item in the builder and on the corresponding student screen**

Several studies (Heffernan et al., 2006; Turner et al., 2005) have been conducted to evaluate the authoring tools in terms of usability and decreased creation time of tutors. We augmented the builder to track how long it takes authors to create an ASSISTment[ii], then invited content creators (undergraduate students and middle school math teachers, all without computer programming experience) to build tutors using the builder. Once we know how many ASSISTments authors have created, we can estimate the amount of content tutoring time created by using the previously established number that students spend about 2 minutes per ASSISTment (Heffernan et al., 2006). This number is averaged from data from thousands of students. This produced a ratio of development time to on-line instruction time of about 40:1, comparing against the literature suggesting a 200:1 ratio (Anderson et al., 1995). The result showed that our pseudo-tutor-based system can reduce the time required to build a single hour of content from 100 to 1000 hours to 10 to 30 hours (Heffernan et al., 2006). This suggests that our method for creating tutoring content is much more cost effectively. We did this by building a tool that reduces both the skills needed to create content as well as the time needed to do so. The determination of whether the ASSISTments created by the content authors produces significant learning is work in progress, however, our subject matter expert was satisfied that the content created was of good quality.

## Reporting

Schools seek to use the yearly MCAS assessments in a data-driven manner to provide regular and ongoing feedback to teachers and students on progress towards instructional objectives. But teachers do not want to wait six months for the state to grade the exams. Teachers and parents also want better feedback than they currently receive. The reporting (Feng & Heffernan, 2007) in the ASSISTments System has been built to identify the difficulties individual students - and the class as a whole – are having. It is intended that teachers will be able to use this detailed feedback to tailor their instruction to focus on the particular difficulties identified by the system.

| Student Name | Elapsed time(hh:mm) | Original Items | | | | Scaffolding + Original Items | | | Most Difficult Learning Standard |
|---|---|---|---|---|---|---|---|---|---|
| | | # Done | % Correct | MCAS Score* | Performance Level | # Done | % Correct | # Hint Req. | |
| Tom* | 4:12 | 90 | 38% | 214 | Warning/Failing-High | 228 | 44% | 233 | N.1.8 Understanding -number -representation |
| Dick* | 4:01 | 98 | 66% | 244 | Pro./Adv. | 158 | 59% | 58 | P.1.8 Understanding -patterns |
| Harry* | 4:07 | 58 | 40% | 219 | Needs improv.-Low | 154 | 38% | 77 | P.7.8 Setting-up-and-solving-equations |

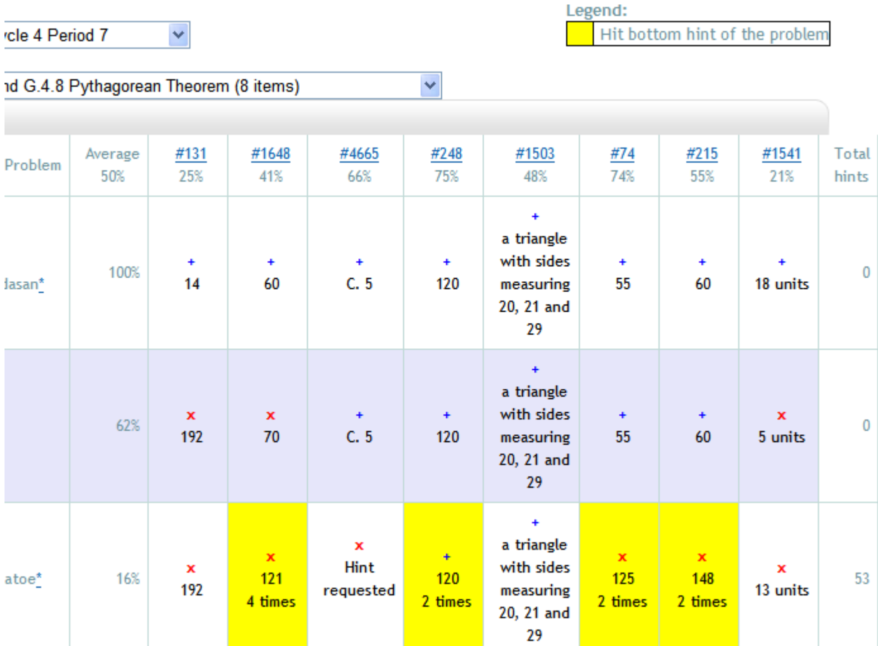
**Figure 3. Grade book report**

[INSERT FIGURE 3 ABOUT HERE: Grade book report]

The "Grade Book", shown in Figure 3, is the ASSISTment report used most frequently by teachers. Each row in the report represents information for one student, including our prediction of his MCAS score based on student response to the original questions. Besides presenting information on the item level, it also summarizes the student's actions in "ASSISTment metrics" which tells more about students' actions besides their performance. For example, it illuminates students' unusual behaviour, such as making far more attempts and requesting more hints than other students in the class. By clicking the link of the skill that the student has the lowest percent correct, the teacher can see what those questions were and what kind of errors the student made. Knowing students' reactions to questions helps teachers to improve their instruction and enable them to correct students' misunderstandings in a straightforward way. Finding out students' difficult knowledge components may also help us improve our item sequencing strategies.

The grade book report gives an overview of a student/a class's performance. Figure 4 shows an item report which shows teachers how students are doing on individual problems. By presenting in different colours and using different tags, the report helps teachers quickly tell if a student answered the question correctly; if not, did they give incorrect answer at their first attempt or they requested for hint, or asked the tutor to break the item into steps. Teachers can tell at a glance which students are asking for too many bottom-out hints (cells are colored in yellow). Teachers can also see what students have answered for each question.

[INSERT FIGURE 4 ABOUT HERE: An item report tells teachers how students are doing on individual problems]




| Problem | Average 50% | #131 25% | #1648 41% | #4665 66% | #248 75% | #1503 48% | #74 74% | #215 55% | #1541 21% | Total hints |
|---|---|---|---|---|---|---|---|---|---|---|
| lasan* | 100% | + 14 | + 60 | + C. 5 | + 120 | + a triangle with sides measuring 20, 21 and 29 | + 55 | + 60 | + 18 units | 0 |
| | 62% | x 192 | x 70 | + C. 5 | + 120 | + a triangle with sides measuring 20, 21 and 29 | + 55 | + 60 | x 5 units | 0 |
| atoe* | 16% | x 192 | x 121 4 times | x Hint requested | + 120 2 times | + a triangle with sides measuring 20, 21 and 29 | x 125 2 times | x 148 2 times | x 13 units | 53 |

**Figure 4.** An item report tells teachers how students are doing on individual problems.

Teachers seem to think highly of the ASSISTment system not only because their students can get instructional assistance in the form of scaffolding questions and hint messages while working on real MCAS items, but also because they can get online, live reports on students' progress while students are using the system in the classroom.

## Content development and management

We are attempting to support the full life cycle of content authoring and management with the tools available in the ASSISTment system. Teachers can create problems with tutoring, map each question to the skills required to solve them, bundle problems together in sequences that students work on, view reports on students' work and use tools to maintain and refine their content over time.

Figure 5 shows how 1) students login, 2) get assignments to do, which then show up such as in the right hand side of Figure 1. Figure 2 also shows that our web-based system allows teachers access to 3) get reports, 4) manage classes, 5) get reports on students, 6a) create, edit and maintain content with the builder, 6b) find their own and others people's content (such as their students' content) 6c-e) bundling that content and assigning it to their students. We even have working reports (step 7) that automatically analyze the results of experiments that randomly assign students to conditions, which is the sort of analysis we need to determine if learning is happening.

[INSERT FIGURE 5 ABOUT HERE: ASSISTments attempt to support the full life cycle of content authoring]

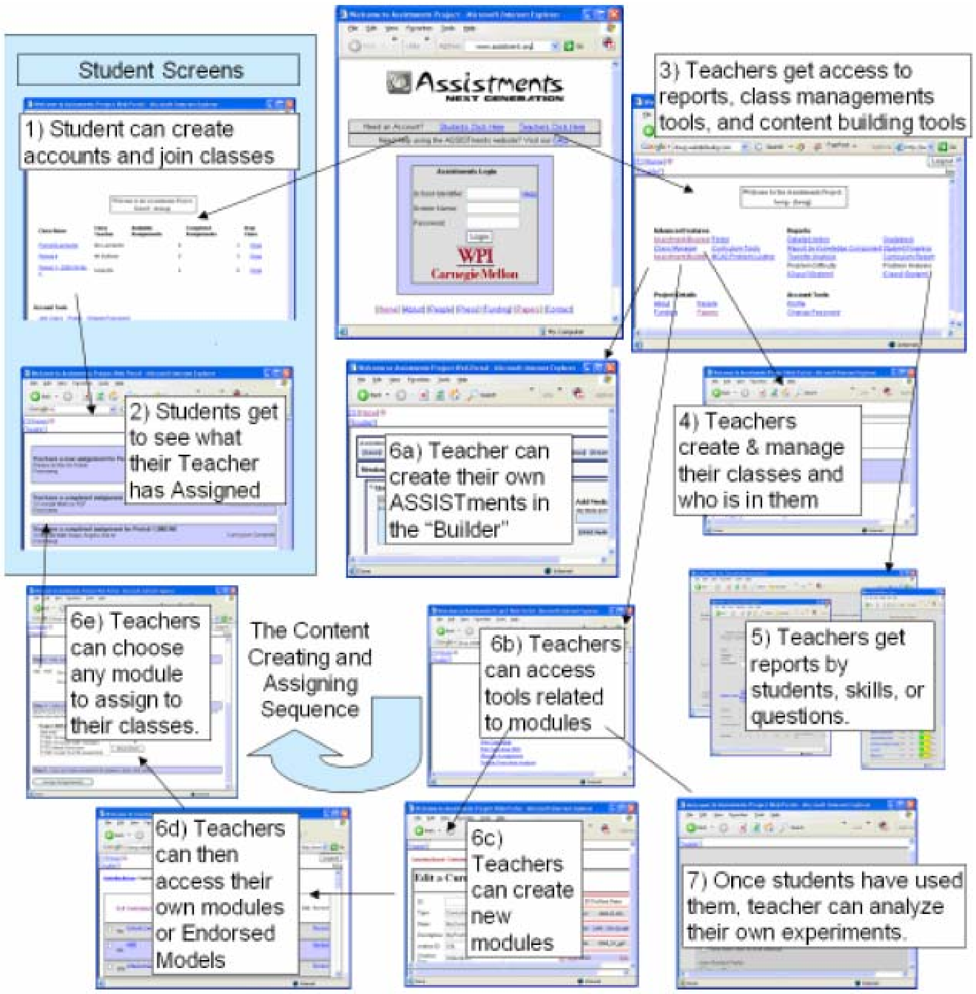



**Figure 5.** ASSISTments attempt to support the full life cycle of content authoring.

## Analyzing Learning effectiveness in ASSISTments

We analyze data within ASSISTments usage to determine whether assistment effectively teaches. For the studies reported by Feng, Heffernan, & Koedinger (2006a, 2006b), we used the ASSISTments system to track student knowledge longitudinally over the course of a schools year, based upon each student using our system about a dozen times during the course of the year. This result confounded learning from the computer system with students learning from their sitting their normal class. To eliminate this confound, Feng, Heffernan, Beck & Koedinger (2008) looked to see if students were reliably learning from their time spent with the *computer in a single day*. We conducted a focused analysis of a subset of items. Items that have the same deep features or knowledge requirements, like approximating square roots, but have different surface features, like cover stories,

were organized into a Group of Learning OPportunity (GLOP). We assessed learning by comparing student performance the first time they were given one item from a GLOP with their performance when they were given more items (also more opportunities) from the same GLOP in the same day. If students tend to perform better on later opportunities of items in a GLOP, it indicates that they may have learned[iii] from the instructional assistance provided on items by the ASSISTments system that they worked on earlier by answering the scaffolding questions or by reading hint messages. Our result suggests that students performed better later in the same computer session on similar skills, which indicates students are learning from using ASSISTments. However, learning is rather uneven across groups of skills. We brought up a variety of hypotheses to explain this phenomenon. We test a few hypotheses and found that the automated approaches we tried were unable to account for the variation. However, human expert judgments were predictive as to which groups of skills were learnable. We found out that students are learning and they learned more from group of items that are more cohesive.

As seen from our student survey results, students complained that being forced into scaffolding questions is time consuming and frustrating. And we were not sure if all of the time we invested into these “fancy” scaffolding questions was worth it. Luckily, the assistment system provides a good platform to run randomized controlled experiments to find out answers like this. Several randomized controlled experiments were designed and carried out to compare the effectiveness of different tutoring strategies. Razzaq & Heffernan (2006) and Razzaq, Heffernan & Lindeman (2007) reported the experiments comparing different tutoring strategies, hints vs. scaffolding. vs. delayed feedback[iv]. And the results showed that in general, scaffolding led to higher averages on a post-test, although it was not statistically significant. And after a closer examine of the effect of math proficiency and the level of feedback on learning, they found out that honored students benefit more from delayed feedback and the regular students did better in the scaffolding condition. One possible explanation is that less proficient students benefit from more interaction and coaching through each step to solve a problem while more proficient students benefit from seeing problems worked out and seeing the big picture. Another study was reported in Razzaq et al. (2005) where the experiment was designed to compare two different tutoring strategies when dealing with proportional reasoning problems. One of the conditions of the experiment involved a student solving two problems like this with scaffolding that first coached them to set up a proportion. The second strategy coached students through the problem but did not use the formal notation of a proportion. Evidence was found that these two different scaffolding strategies seem to have different rates of learning. However, the fact that one strategy seems better than the other is not the point. The point is that it is a future goal for the ASSISTments system to do this sort of analysis automatically for content creators, and based on the research results pick the most suitable tutoring for the learners.

## Student Modeling in ASSISTments

The ASSISTments system has two assessment goals: predicting student performance on end-of-year accountability exams; and cognitively assess student knowledge to help teachers target next instructional steps. These goals are complicated by two features of the system: assessment is ongoing throughout the school year as student proficiency develops; and the ASSISTments system itself as tutoring system changes student proficiency. Nevertheless, prediction using simple student models and a number of “assistance metrics” (summaries of hint-seeking behavior, time spent on questions correctly vs. incorrectly answered, etc.) can be almost optimally effective at predicting end-of-year exam scores. Statistical uncertainty in teacher feedback reports based on more-detailed student models is sometimes surprisingly low, but even in cases where the per-student uncertainty is high, reports aggregated over groups of students can be quite reliable. These ideas will be considered both for the ASSISTments system and in the broader context of online assessment and learning systems.

## Predicting student end-of-year exam score

The first assessment goal for ASSISTments is to make a prediction of student end-of-year exam score. We have reported the results of several studies where we use data collected via ASSISTments to try to predict state test scores required of all students in the state. We will now review of few of those results. In all of these studies we report compared students actual 2005 MCAS test scores with the predicted to get calculate the Mean Absolute Deviation (MAD) which was then used as the measure to evaluate the student models. Does providing assistance hurt the accuracy of the assessment? Surprisingly, these studies reported that the assistance provided actually improves the assessment. The idea is that by seeing how much help students needed allows a more sensitive measure of student knowledge then just whether they got a question correct.

### *Dynamic testing*

Much work has been done in the past 10 years or so on developing “online testing metrics” for dynamic testing (or dynamic assessment) (Grigorenko and Sternberg, 1998) to supplement accuracy data (wrong/right scores) for characterizing student proficiency. Researchers have been interested in trying to get more assessment value by comparing traditional assessment (students getting an item marked wrong or even getting partial credit) with a measure that shows how much help they needed. Brown, Bryant, and Campione (1983) compared traditional testing paradigms against a dynamic testing paradigm. Grigorenko & Sternberg (1998) reviewed relevant literature on the topic and expressed enthusiasm for the idea. In the dynamic testing paradigm a student would be presented with an item and when the student appeared to not be making progress, would be given a prewritten hint. If the student was still not making progress, another prewritten hint was presented and the process was repeated. In this study they wanted to predict learning gains between pretest and posttest. They found that static testing prediction did not correlate ($R = 0.45$) with student learning data as well as the “dynamic testing” did ($R = 0.60$). Brown et al. (1983) suggested that this method could be effectively done by computer, but, as far as we know, their work was not continued. Luckily, the ASSISTment system provides an ideal test bed as it already provides a set of hints to students. So it is a natural way to extend and test this idea and see if we can replicate their finding of ASSISTment-style measures being better assessors. Our hypothesis is that we can achieve more accurate assessment by not only using data on whether students get test items right or wrong, but by also using data on the effort required for students to learn how to solve a test item.

We continued the dynamic testing approach (Feng, Heffernan & Koedinger, 2006a, 2006b; Feng et al., 2008) and developed a group of “assistance” metrics that measures students’ accuracy (correct or incorrect responses), speed (how many seconds a student needs to solve a problem), attempts (how many attempts a student made to finally get the correct answer) and help-seeking behavior (how often a student asks for hints). We computed the metrics from our log data and found out that all these metrics are reliably correlated with student real MCAS test scores. Then we built different students models to predict end-of-year MCAS scores. Our goal was to see if we can reliably predict students’ test scores and to evaluate how well on-line use of the ASSISTments System can help in the prediction.

A number of different models were compared for measuring student knowledge during ASSISTments use. The key contrast of interest is between a static model that mimics paper practice tests by scoring students as either correct or incorrect on each main question, with a dynamic assessment model that leverages the assistance metrics to take into account the amount of assistance students need before they get an item correct. In the data set we collected during the school year of 2004-2005, the predicted score from the static model correlated with the MCAS test scores, with an R-value of 0.731 and the dynamic assistance model correlated with an R-value of 0.865, reliably higher than the correlation between the static prediction and MCAS scores. Thus, there is some evidence that showing the ASSISTments system is doing a good job at predicting student math proficiency. Additionally we can improve our prediction of MCAS score further by including the assistance metrics in a dynamic student model.

It is suspected that a better job of predicting MCAS scores could be done if students could be encouraged to take the system seriously and reduce “gaming behavior”. One way to reduce gaming is to detect it and then to notify the teacher's reporting session with evidence that the teacher can use to approach the student. Our preliminary work on gaming detection was presented in Walonoski & Heffernan (2006). It is assumed that teacher intervention will lead to reduced gaming behavior, and thereby more accurate assessment, and higher learning. Adding visual feedback, as one ongoing work in the ASSISTment system does, aims to help teachers quickly detect gaming behaviors.

### *Tracking student performance longitudinally*

In Razzaq et al. (2005) and Feng, Heffernan, Beck & Koedinger (2008), we reported results that suggested students were learning directly during the assisting in ASSISTments. We did this by looking at groups of items that had the same skills and looked to see if performance later in the class period was associated with high performance. The gain score over all of the learning opportunity pairs suggests that students were learning in the system. In this section, instead of discussing within-system learning, we focus on tracking student progress that results from both classroom instruction and ASSISTments tutoring over a long period of time. To investigate this question, we did a longitudinal analysis (Singer & Willett, 2003, Fitzmaurice, Laird & Ware, 2004) by fitting a mixed-effect model on the ASSISTment data to investigate if learning happens over time. We gradually introduced factors such as what school they are in, who their teacher is, or which class they are from into our models. By doing so, we attempt to provide an answer to the question of what factors impact (or are correlated with) students’ learning rate.

During the school year of 2004-2005, there were approximately 650 students using the system, with each student coming to the computer lab about 7 times, there was a table with 4550 rows, one row for each student for each day, with an average percent correct which itself is averaged over about 15 MCAS items done on a given day. In Figure 6, average student performance is plotted versus time. The y-axis is the average percent correct on the original item (student performance on the scaffolding questions is ignored in this analysis) in a given class. The x-axis represents time, where data is bunched together into months, so some students who came to the lab twice in a month will have their numbers averaged. The fact that most of the class trajectories are generally rising suggests that most classes are learning between months. The result of statistical modeling confirms what we saw in the plot. Our fitted longitudinal model ended up with a statistically significant learning slope which indicates that student performance was reliably increasing during the school year. Additionally, our model was able to detect different rates of learning at different schools, but not among different teachers and classes.

[INSERT FIGURE 6 ABOUT HERE: Average student performance plotted over time]

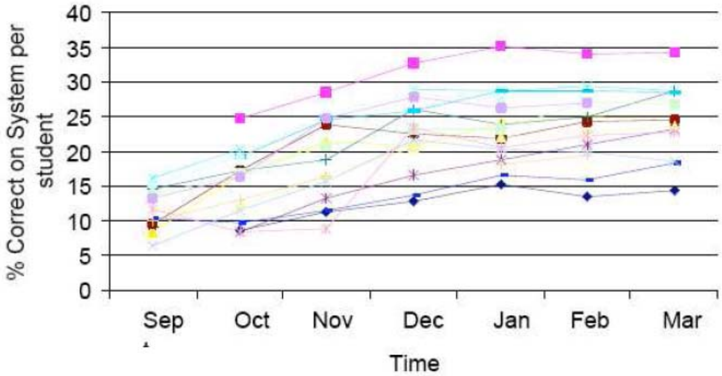



**Figure 6. Average student performance plotted over time**

Given that this is the first year of the ASSISTments project, new content is created each month, which introduces a potential confounder of item difficulty. It could be that some very hard items were selected to give to students in September, and students are not really learning but are being tested on easier items. In the future, this confound will be eliminated by sampling items randomly. Adding automated applied longitudinal data analysis is currently being pursued.

More work is needed to build models to better be able to detect differences between teachers' effects on the learning rates of students that presumably exist. Besides, other factors will be investigated about their possible impact on students' learning over time. Information from student profiles such as gender, race and ethnicity, special education status, free-lunch status, etc., is in our consideration. During this analysis, we noticed the fact that generally speaking, groups with higher estimated initial scores showed lower rates of learning. Our preliminary speculation on this fact is that 1) this may be attributed to the "ceiling effect": it is hard for top students to make fast progress; 2) good students were assigned to Algebra class and learning content that won't be tested until $10^{th}$ grade and won't appear in the ASSISTment system. Further investigation needs to be done to explain this phenomenon.

## Modeling fine grained student knowledge

### *Cognitive models development and skill mapping*

The ASSISTments approach is task-centric (development of main questions and scaffolding materials starts from released state exam items); but it directly attributes individual differences to unobservable skills or other latent variables. Most large standardized tests (like the math-subtest of the Graduate Record Examination (GRE)) are what psychometricians call "unidimensional" in that they are analyzed as if all the questions are tapping a single underlying knowledge component (i.e., skill). It is this assumption of unidimensionality that makes computer adaptive testing possible for the GRE. However, cognitive scientists such as Anderson & Lebiere (1998), believe that students are learning individual skills, and might learn one skill but not another.

In April, 2005, we invited educational researchers and subject manner experts to conduct cognitive task analysis over the released state test items. They set out to make up skills and tag the entire existing 8th grade MCAS items with these skills. There were about 300 released test item to code. Because we wanted to be able to track learning between items, we wanted to come up with a number of skills that were somewhat fine-grained but not too fine-grained such that each item had a different skill. We therefore imposed upon our subject-matter expert that no one item would be tagged with more than 3 skills. The subject manner expert was free to make up whatever skills she thought appropriate. Although we have English names for the skills, those names are just a handy tag; the real meaning of a skill must be divined by the questions with which it is associated. We ended up with a cognitive model of 106 skills which we refer to as WPI-106.

One aspect of the project is that the ASSISTments system must serve a variety of stakeholders, and not all of them need or want reports at the same level of granularity. Therefore, even though WPI-106 may be closer to optimal for providing teacher feedback, we built other cognitive models at coarser grain size[v]. We used the fine-grained model to guide us to create the coarse-grained models. We decided to use the same 5 strands that both the National Council of Teachers of Mathematics uses, as well as the Massachusetts Department of Education. These categories are named 1) “Patterns, Relations and Algebra”, 2) “Geometry”, 3) “Data Analysis, Statistics and Probability”, 4) “Number Sense and Operations” and 5) “Measurement”. The Massachusetts Department of Education actually tags each item with exactly one of the 5 categories, but our mapping was not necessarily the same as the states’. Furthermore, we allowed multi-mapping, i.e. allow an item to be tagged with more than one skill. A middle-level model of 39 skills was also derived by nesting a group of skills from WPI-106 into the each one of the 39 learning standards from Massachusetts Curriculum Framework.

Table 1 shows the hierarchal nature of the relationship among the three models of different grain size. We think we will be able to track student knowledge in the fine grained model, partly because students usually finished many problems along the year (on average 100+ main problems, 200+ scaffolding questions) and partly because the strategy of scaffolding questions gives us the identifiablity. As mentioned before, when creating scaffolding questions for the tutoring session, the authors try to focus on one skill at a time step by step, which makes the scaffolding questions as *cognitive diagnostic assessment* on top of tutoring and therefore student response to scaffolding questions can be used to determine what are the skills students have mastered or have difficulty on.

[INSERT TABLE 1 ABOUT HERE: Cognitive models in hierarchically nested structure]

**Table 1. Cognitive models in a hierarchically nested structure**

| WPI-106 | MCAS-39 | MCAS-5 |
|---|---|---|
| Inequality-solving<br>Equation-Solving<br>Equation-concept | setting-up-and-solving-equations | Patterns-Relations-Algebra |
| … | … | |
| Plot Graph | modeling-covariation | |
| X-Y-Graph<br>Slope | understanding-line-slope-concept | |
| … | … | … |
| Congruence<br>Similar Triangles | understanding-and-applying-congruence-and-similarity | Geometry |
| … | | |
| … | … | |
| Perimeter<br>Circumference<br>Area | using-measurement-formulas-and-techniques | Perimeter |
| … | | |
| … | … | |

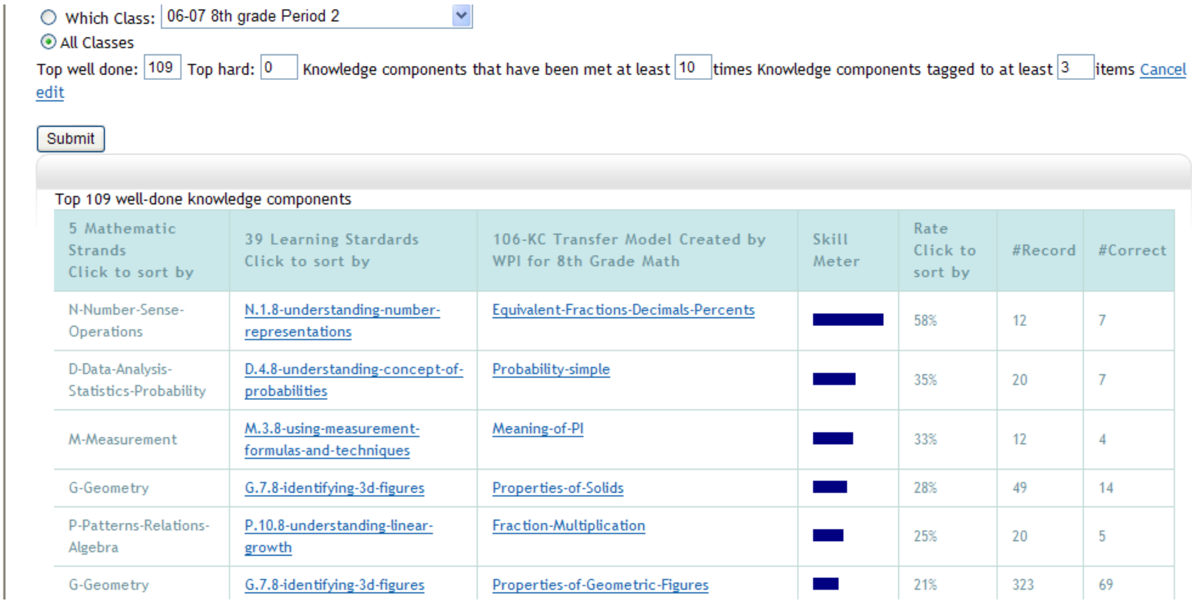
A secondary purpose of the ASSISTments builder was to aid the mapping between skills and the questions. As they are building content, the authors use the builder to tag certain skills to specific problems to indicate that a problem requires knowledge of that skill.

### *Inference of skill level and reporting to teachers*

Mapping between skills and problems allows our reporting system to track student knowledge over time using longitudinal data analysis techniques (Singer & Willett, 2003; Fitzmaurice, Laird, & Ware, 2004) and make inference of student proficiency level on each skills based on their performance on the problems. In ASSISTments, the inference of student proficiency level is rather simple (Feng, Heffernan, Mani & Heffernan, 2006). Students get full credit for a skill when they correctly answer the questions tagged with the skill. In the case of a wrong answer to a question tagged with multiple skills, the system relies on response to scaffolding questions (typically tagged with only one skill) to determine which skill "to blame" (i.e., the cause of the wrong answer to the main question). If no scaffolding questions available, the most difficult skill will be blamed. Thus, connection between proficiencies and tasks is relatively loose and informal.

Turning to teacher feedback, Figure 7 shows the skill analysis report that informs teachers about the knowledge status of selected classes. Skills (labelled as Knowledge components in the report) are ranked according to their correct rate—labelled “Rate” and displayed as “Skill Meter”—which is the percent correct rate at the items tagged with that skill model. Skills in the reports are organized hierarchically in three levels of grain sizes. Links in the column entitled “39 learning standards” will lead to the definition of a particular learning standard in Mathematics Curriculum Framework. Clicking the name of a skill in WPI-106 (labelled as “106-KC Transfer Model”), teachers will be redirected to another page showing the items tagged with that skill. In the new page, teachers are able to see the question text of each item and continue to preview or analyze the item if they want to know more about it. By presenting such a report, we hope we can help teachers to decide which skill and items should be focused on to maximize the gain of students' scores at a class level when instructional time is limited.

[INSERT FIGURE 7 ABOUT HERE: Class level skill analysis report]



Which Class: 06-07 8th grade Period 2

All Classes

Top well done: 109 Top hard: 0 Knowledge components that have been met at least 10 times Knowledge components tagged to at least 3 items Cancel edit

Submit

Top 109 well-done knowledge components

| 5 Mathematic Strands Click to sort by | 39 Learning Stardards Click to sort by | 106-KC Transfer Model Created by WPI for 8th Grade Math | Skill Meter | Rate Click to sort by | #Record | #Correct |
|---|---|---|---|---|---|---|
| N-Number-Sense-Operations | N.1.8-understanding-number-representations | Equivalent-Fractions-Decimals-Percents | | 58% | 12 | 7 |
| D-Data-Analysis-Statistics-Probability | D.4.8-understanding-concept-of-probabilities | Probability-simple | | 35% | 20 | 7 |
| M-Measurement | M.3.8-using-measurement-formulas-and-techniques | Meaning-of-PI | | 33% | 12 | 4 |
| G-Geometry | G.7.8-identifying-3d-figures | Properties-of-Solids | | 28% | 49 | 14 |
| P-Patterns-Relations-Algebra | P.10.8-understanding-linear-growth | Fraction-Multiplication | | 25% | 20 | 5 |
| G-Geometry | G.7.8-identifying-3d-figures | Properties-of-Geometric-Figures | | 21% | 323 | 69 |

**Figure 7. class level skill analysis report**

### *The effect of model granularity on student performance prediction*

We then engaged in an effort to investigate if we can do a better job of end-of-year standardized test by modeling individual skills in a finer grain size (Feng, Heffernan, Mani & Heffernan, 2006; Pardos, Feng, Heffernan & Heffernan, 2007; Pardos, Heffernan, Anderson & Heffernan, 2006). We consider the 4 different models, unidimensional, 5 skills, 39 skills, and 106 skills. We fit mixed-effects logistic model to track student knowledge change longitudinally and then made a prediction of how a student will perform on each skill in the MCAS test[vi] at the end of the year. The measure of model performance is the accuracy of the predicted MCAS test score based on the assessed skills of the students. We found out that the WPI-106 model is superior in terms of prediction accuracy of 2005 MCAS test, followed by the model with 39 skills. And including student's response to scaffolding questions helped building a better prediction model, too. We also tried to add item difficulty parameter obtained from a Rasch model (van der Linden & Hamilton, 1997) into the mixed-effects logistic model as an additional covariate to account for the fact that questions tagged with the same skill may vary on difficulty, but the result suggested that item difficulty parameter does not help on top of skill tracking (Feng & Heffernan, 2007a).

## Conclusion and General Implications

In this chapter, we address the student modeling challenge in the ASSISTments system, a web-based e-learning and e-assessment system. We concentrate on the assessment ability of the system. Some evidence was presented that the online assessment system did a better job of predicting student knowledge by being able to take into consideration how much tutoring assistance was needed. Promising evidence was also found that the online system was able to track students' learning during a year well. Furthermore, we showed how individual skills were modeled in the ASSISTment system and being used to give feedback to teachers.

Traditional assessment usually focuses on students' responses to test items and whether they are answered correctly or incorrectly. It ignores all other student behaviors during the test (e.g., response time). However, in this work, we take advantage of a computer-based tutoring system to collect extensive information while students interact with the system. Our results showed that the assistance model leads to significantly better predictions than the model that is based on the assessment results alone. Not only is it possible to get reliable test information while "teaching on the test", data from the teaching process actually improves reliability.

Currently we are beginning to focus statistical modeling work on improving student knowledge modeling in the ASSISTments system. For example, Cen, Koedinger & Junker (2006) model learning curves using ideas of Draney, Pirolli and Wilson (1995). Another approach as elaborated in the Evidence Centered Design (ECD) (Mislevy, Steinburg, & Almond, 2003) framework gives special attention to the role of probability-based reasoning in accumulating evidence across task performances, in terms of belief about unobservable variables that characterize the knowledge, skills, and/or abilities of students. Another approach combines the knowledge tracing algorithm of Corbett, Anderson & O'Brien (1995) with Bayes Net (DINA) models (Junker & Sijtsma, 2001). The key criterion in determining which approach to pursue will be model fit and interpretability.

The more general implication from this research suggests that continuous assessment systems are possible to build and that they can be quite accurate at helping schools get information on their students. This result is important because it provides evidence that reliable assessment and instructional assistance can be effectively blended. These results with the ASSISTment system open up a the possibility of a completely different approach to assessment that is contentious in nature in suggesting students may not need to spend any time on formal paper and pencil tests. Many states are moving towards adopting "*value added*" assessments, so that they can track the *value added* by teachers and schools. *Value added* is possibly because you have year to year state assessments so you can see the average learning gain for students per year, and attribute those gains to teachers and schools. Tennessee is even paying teachers differently based upon a teacher's averaged gain score[vii]. Such systems could benefit from data that is collected every two weeks, instead of once a year, thereby allowing schools to more quickly figure out *what works* at increasing student learning. Because the ASSISTment system teaches while it assesses, it makes the testing more politically palatable. In fact, this paper showed that because the system teaches while it assesses, its does a better job of assessing (if you hold the number of items done constant, instead of time). One might also be concerned that using the ASSISTment system may take longer than taking a paper practice test. However, unlike paper tests, the ASSISTment system is contributing to instruction (Razzaq et al., 2005; Feng, Heffernan, Beck & Koedinger, 2008). While every minute spent on a paper test takes away a minute of instruction, every minute on the ASSISTment system contributes to instruction. We end with a tantalizing question: Are we likely to see states move from a test that happens once a year, to an assessment tracking system that offers continuous assessment (Computing Research Association, 2005) every few weeks? While more research is warranted, our results suggest that perhaps the answer should be yes.

## Acknowledgement

This research was made possible by the U.S. Department of Education, Institute of Education Science (IES) grants, "Effective Mathematics Education Research" program grant #R305K03140 and "Making Longitudinal Web-based Assessments Give Cognitively Diagnostic Reports to Teachers, Parents, & Students while Employing Mastery learning" program grant #R305A070440, the Office of Naval Research grant #N00014-03-1-0221, NSF CAREER award to Neil Heffernan, and the Spencer Foundation. All the opinions, findings, and conclusions expressed in this article are those of the authors, and do not reflect the views of any of the funders. We would like to acknowledge Joseph Beck for his insightful comments on this research.

## References

[1] 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.

[2] Anderson, J. R. & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum.

[3] Aleven, V.A.W.M.M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26(2), pp.149-179.

[4] Anozie, N.O.& Junker, B.W. (2006). *Predicting end-of-year accountability assessment scores from monthly student records in an online tutoring system.* American Association for Artificial Intelligence Workshop on Educational Data Mining (AAAI-06), July 17, 2006, Boston, MA.

[5] Ayers, E. & Junker, B.W. (2006). *Do skills combine additively to predict task difficulty in eighth-grade mathematics?* American Association for Artificial Intelligence Workshop on Educational Data Mining (AAAI-06), July 17, 2006, Boston, MA.

[6] Beck, J.E. (2006). Using learning decomposition to analyze student fluency development. *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems*. Jhongli, Taiwan. pp. 21-28.

[7] Brown, A. L., Bryant, N.R., & Campione, J. C. (1983). Preschool children's learning and transfer of matrices problems: Potential for improvement. Paper presented at the Society for Research in Child Development meetings, Detroit.

[8] Campione, J.C., Brown, A.L., & Bryant, N.R. (1985). Individual differences in learning and memory. In R.J. Sternberg (Ed.). *Human abilities: An information-processing approach*, 103–126. New York: W.H. Freeman.

[9] Cen, H., Koedinger, K., & Junker, B. (2006). Learning factor analysis – A general method for cognitive model evaluation and improvement. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems.* Springer-Verlag: Berlin. pp. 164-175.

[10] Corbett, A. T., Anderson, J. R., & O'Brien, A. T. (1995) Student modeling in the ACT programming tutor. Chapter 2 in P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively Diagnostic Assessment* Hillsdale, NJ: Erlbaum.

[11] Corbett, A. T., Koedinger, K. R.,& Hadley,W. H. (2001). Cognitive Tutors: From the research classroom to all classrooms. In Goodman, P. S. (Ed.) *Technology Enhanced Learning: Opportunities for Change.*Mahwah, NJ: Lawrence Erlbaum Associates.

[12] Computing Research Association. (2005). Cyberinfrastructure for Education and Learning for the Future: a Vision and Research Agenda. Final report of Cyberlearning Workshop Series workshops held Fall 2004 - Spring 2005 by the Computing Research Association (CRA) and the International Society of the Learning Sciences (ISLS).

[13] Draney, K. L., Pirolli, P., & Wilson, M. (1995). A measurement model for a complex cognitive skill. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively Diagnostic Assessment.* Hillsdale, NJ: Erlbaum.

[14] Feng, M., Heffernan, N. T., & Koedinger, K. R. (2006a). Addressing the testing challenge with a web based E-assessment system that tutors as it assesses. *Proceedings of the 15th Annual World Wide Web Conference.* pp. 307-316. ACM Press: New York.

[15] Feng, M., Heffernan, N. T., & Koedinger, K. R. (2006b). Predicting state test scores better with intelligent tutoring systems: developing metrics to measure assistance required. In Ikeda, Ashley & Chan (Eds.) *Proceedings of the Eighth International Conference on Intelligent Tutoring Systems.* pp 31–40. Springer-Verlag: Berlin.

[16] Feng, M., Heffernan, N., Mani, M., & Heffernan, C. (2006). Using mixed effects modeling to compare different grain-sized skill models. In Beck, J., Aimeur, E., & Barnes, T. (Eds). *Educational Data Mining: Papers from the AAAI Workshop*. Menlo Park, CA: AAAI Press. pp. 57-66. Technical Report WS-06-05..

[17] Feng, M., & Heffernan, N.T. (2006). Informing Teachers Live about Student Learning: Reporting in the Assistment System. *Technology, Instruction, Cognition, and Learning Journal*, Vol. 3. Philadelphia, PA: Old City Publishing.

[18] Feng, M. & Heffernan, N. T. (2007a). Assessing Students' Performance: Item Difficulty Parameter vs. Skill Learning Tracking. Paper presented at the National Council on Educational Measurement 2007 Annual Conference, Chicago.

[19] Feng, M. & Heffernan, N. (2007b). Towards Live Informing and Automatic Analyzing of Student Learning: Reporting in ASSISTment System. *Journal of Interactive Learning Research*. 18(2), pp. 207-230. Chesapeake, VA: AACE.

[20] Feng, M., Beck, J,. Heffernan, N. & Koedinger, K. (2008). Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test? In Beck & Baker (Eds.). *Proceedings of the 1st International Conference on Education Data Mining.* pp.107-116.

[21] Feng, M., Heffernan, N., Beck, J, & Koedinger, K. (2008). Can we predict which groups of questions students will learn from? In Beck & Baker (Eds.). *Proceedings of the 1st International Conference on Education Data Mining.* pp.218-225.

[22] Fitzmaurice, G., Laird, N., & Ware, J. (2004). *Applied Longitudinal Analysis*. Hoboken, NJ:Wiley & Sons.

[23] Grigorenko, E. L. and Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124,* 75–111.

[24] Junker, B.W. & Sijtsma K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25,* 258–272.

[25] Heffernan N.T., Turner T.E., Lourenco A.L.N., Macasek M.A., Nuzzo-Jones G., & Koedinger K.R. (2006). The ASSISTment Builder: Towards an Analysis of Cost Effectiveness of ITS creation. In Sutcliffe, G., & Goebel R. (Eds.), *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference*. pp. 515-520.

[26] 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.

[27] Koedinger, K. R., Aleven, V., Heffernan. N. T., McLaren, B. & Hockenberry, M. (2004). Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems,* pp. 162-173, Maceio, Brazil.

[28] Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives,* 1, 3-67.

[29] Olson, L. (2005). Special report: testing takes off. *Education Week*, November 30, 2005, pp. 10–14. Also available on-line from http://www.edweek.org/media/13testing.pdf

[30] Pardos, Z. A., Feng, M., Heffernan, N. T. & Heffernan, C. L. (2007). Analyzing Fine-Grained Skill Models Using Bayesian and Mixed Effect Methods. In Luckin & Koedinger (Eds.) *Proceedings of the 13th Conference on Artificial Intelligence in Education* (AIED 2007). pp. 626-628.

[31] Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2006). *Using Fine Grained Skill Models to Fit Student Performance with Bayesian Networks.* Workshop in Educational Data Mining held at the Eighth International Conference on Intelligent Tutoring Systems. Taiwan. 2006.

[32] Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D.Wright. Chicago: The University of Chicago Press.

[33] 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.* pp 555–562. Amsterdam: ISO Press.

[34] Razzaq, L., & Heffernan, N.T. (2006). Scaffolding vs. hints in the Assistment System. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. Berlin, Springer-Verlag. pp. 635-644. 2006.

[35] 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.* Amsterdam, Netherlands: IOS Press.

[36] Razzaq, L., Feng, M., Heffernan, N. T., Koedinger, K. R., Junker, B., Nuzzo-Jones, G., Macasek, N., Rasmussen, K. P., Turner, T. E. & Walonoski, J. (2007). A web-based authoring tool for intelligent tutors: blending assessment and instructional assistance. In Nedjah, N., et al. (Eds). *Intelligent Educational Machines*. Intelligent Systems Engineering Book Series.

[37] Rothman, S. (2001). 2001 MCAS Reporting Workshop: The second generation of MCAS results. Massachusetts Department of Education. Downloaded November 2006 from http://www.doe.mass.edu/mcas/2001/news/reporting wkshp.pps.

[38] Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21, 1-29.

[39] Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and occurrence.* Oxford University Press, New York.

[40] Turner, T. E., Macasek, M. A., Nuzzo-Jones, G., Heffernan, N. T. (2005). The ASSISTment Builder: A Rapid Development Tool for ITS. In *Proceedings of the 12th Annual Conference on Artificial Intelligence in Education*, pp. 929-931.

[41] van der Linden, W. J. & Hambleton, R. K. (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.

[42] 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), 1-47.

[43] Walonoski, J., Heffernan, N.T. (2006). Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. Springer-Verlag: Berlin. pp. 382-391. 2006

Websites:
http://www.assistment.org
http://www.learnlab.org
http://www.educationaldatamining.org

---

[i] The term “Assistment” was coined by Kenneth Koedinger and blends Assessment and Assisting.

[ii] This does ignore the time it takes authors to plan the ASSISTment, work with their subject-matter expert, and any time spent making images and animated gifs. All of this time can be substantial, so we cannot claim to have tracked all time associated with creating content.

[iii] There is controversy over whether same-day learning opportunities should be used as evidence of learning. For example, Beck (2006) thought repeated trials were not indicative of learning, but just retrievals from short term memory.

[iv] Delayed feedback is similar to worked-out examples (Renkl, 1997) in that it shows the solution to a problem all at once. The difference is that students are administered problems first, but get no feedback until they complete a problem set. After finishing each problem, students will be told “we will give feedback at the end of the problem set”.

[v] As the ASSISTment system is considered in multiple states and other jurisdictions, additional transfer models will be needed, that are aligned to those states’ learning standards.

[vi] All items in the MCAS tests are tagged in all the four models before this analysis by our subject manner expert.

[vii] http://www.shearonforschools.com/TVAAS_index.html