Tech., Inst., Cognition and Learning, Vol. 3, pp. 00-00
Reprints available directly from the publisher
Photocopying permitted by license only

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

Informing Teachers Live about Student Learning: Reporting in the Assistment System

MINGYU FENG*, NEIL T. HEFFERNAN

Department of Computer Science, Worcester Polytechnic Institute
Worcester, MA 01609

Limited classroom time available in middle school mathematics classes forces teachers to choose between assisting students' development and assessing students' abilities. To help teachers make better use of their time, we are integrating assistance and assessment by utilizing a web-based system ("Assistment") that will offer instruction to students while providing a more detailed evaluation of their abilities to the teacher than is possible under current approaches (refer to [7] for more details about the Assistment system). We describe the reports designed and implemented to provide real time reporting to teachers in their classrooms. This reporting system is robost enough to support the 800 students currntly using our system.

Keywords: Author to provide keywords

INTRODUCTION

MCAS (Massachusetts Comprehensive Assessment System) is a graduation requirement in which all students educated with public funds in the tested grades are required to participate. This is a high-stakes standards-based test filled with challenging multi-step problems that tap a variety of different mathematical concepts, making it nearly impossible to prepare to this test in a shallow manner. In 2004, over 20% of students failed the 10th

*Corresponding author: mfeng|nth@cs.wpi.edu; (508) 831-5569

grade math test, and 2,582 students, representing 4% of 12th graders were denied a diploma for not passing the test (Massachusetts Dept. of Education, 2003). Because students are more likely to fail the mathematics portion of the test, educators are focusing efforts on mathematics.

Given the limited classroom time available in middle school mathematics classes, teachers must choose between time spent assisting students' development and time spent assessing students abilities. To help resolve this dilemma, we are integrating assistance and assessment by utilizing a web-based system ("Assistment") supported by the U.S. Department of Education. The Assistments system offers instruction to students while providing a more detailed evaluation of their abilities to the teacher than is possible under current approaches. Each assistment consists of an *original item* and a list of *scaffolding questions*[1] given only to students who have given wrong answers to original items. By providing instructional assistance during assessment, teachers justify having students spend time using our system. By breaking the original problems into scaffolding questions plus *hint messages* and *buggy messages*, the Assistment system gives more focused instruction than that provided by online multiple-choice systems. Also the scaffolding questions enable us to assess individual knowledge components instead of only overall performance. Our supporting website "www.assistment.org" has been running for half a year, providing 75 assistments and is being used by 9 teachers and about 1000 students.

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. However, teachers need feedback more often than once a year and they 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. While the number of mathematics skills and concepts that a student needs to acquire is on the order of hundreds, the feedback on the MCAS is broken down into only 5 mathematical reporting categories, known as "Strands". One principal requests more efforts on Geometry and Measurement because his students scored poorly in those areas (receiving 38% and 36% correct compared to over 41+% correct in the three other categories). Additionally, individual student's response to each question is also given in the feedback. However, a detailed analysis of state tests in Texas concluded that such topic reporting is not reliable because items are not equated for difficulty

[1]We use the term scaffolding question because they are like scaffolding that will help students solve the problem (and can "faded" later) so the scaffolds are meant to scaffold their learning. [2]

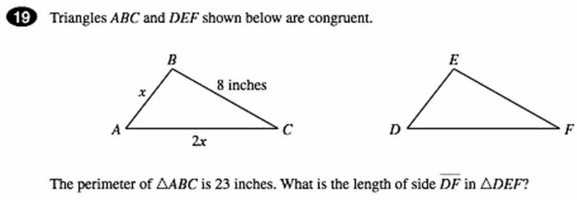


Figure 1
Item 19 from 2003 MCAS.

within these areas [3]. To get some intuition on why this is the case, the reader is encouraged to try item 19 from the 2003 MCAS shown in Figure 1. Then ask yourself "What is the most important thing that makes this item difficult?" Clearly, this item includes elements from four of the 5 "strands" (only missing "Data Analysis, Statistics and Probability"). They are Algebra, Geometry (for its use of congruence), Number Sense (for doing the arithmetic operations), or Measurement (for the use of perimeter). Ignoring this obvious overlap, the state chose just one strand, Geometry, to classify the item. As shown below, the question of tagging items to learning standards is very important because teachers, principals and superintendents are all being told to be "data-driven" and use the MCAS reports to inform their instruction and help their students. As a teacher has said "It does affect reports... because then the state sends reports that say that your kids got this problem wrong so they're bad in geometry-and you have no idea, well you don't know what it really is- whether it's algebra, measurement/perimeter, or geometry."

There are several reasons for this poor reporting: 1) the reasonable desire to give problems that tap-multiple knowledge components, 2) a student's response to paper and pencil tests alone is not sufficient to determine what knowledge components to credit or blame, and 3) some problem knowledge components involved in decomposing and recomposing multi-step problems need to be modelled, yet are currently poorly understood by cognitive science. Accordingly, a teacher cannot trust that putting more effort on a particular low scoring area will pay off in the next round of testing.

In the Assistment Project, we have made an effort to give quick (even live) reports to teachers based on continuous data from their students, including more detailed analysis of students' learning and knowledge status. We have defined a finer-grained transfer model and applied it in our reporting system to address students' weaknesses on specific concepts or skills and to infer what students know and are learning. We believe this makes it easier for a teacher to make data-driven changes in the classroom.

The rest of the paper is organized as follows: The first section introduces the system infrastructure. We describe the MCAS curriculum frameworks and the hierarchical transfer model developed at Worcester Polytechnic Institute (WPI) in the second section. The third section shows an improved reporting system informing teachers in a more efficient and instructive way. It also shows how teachers are helped to analyse problems. Initial results on automating student learning analysis are also presented.

1. DATA COLLECTION

The Assistment system is deployed on the Internet, wherein students open a web browser and login in to work on problems. A Java-based runtime system [5] posts each student's actions (other than mouse movements) to a message server as an xml message that includes timestamp, student ID, problem ID, student's action type (attempt problem or request help), student's input and response. Messages are first queued and then stored in the database server at WPI. Students' progress is remembered in progress files enabling them resume their work.

MySQL database server was used as the database server before switching to Oracle when records increased to 1 million in just half a year. Currently, the database uses 30 tables; one for storing log data, 3 for tracking users, 6 for assistments, curricula, class assignments and students' progress, 5 for storing paper and pencil tests results, about 5 for a transfer model (transfer model structure, knowledge components and mappings between knowledge components and questions). Other database objects (e.g., views, indices, stored procedures) help organize data and promote reporting performance.

2. TRANSFER MODEL

A transfer model [4] is a cognitive model that contains a group of knowledge components and maps existing questions (original items and scaffolding questions) to knowledge components. It also indicates the number of times a particular knowledge component has been applied for a given question. It is called a "transfer model" since the intent is to predict learning and knowledge transfer. Transfer models are useful in selecting the next problem. The next section shows that transfer models are important for quality reporting.

2.1 Massachusetts Curriculum Frameworks

The Massachusetts Curriculum Frameworks (MCAS) breaks 5 strands (Patterns, Relations, and Algebra; Geometry; Data Analysis, Statistics, and Probability; Measurement; Number Sense and Operations) into 39 individual "learning standards" for 8th grade math and tags each item with one of the 39 standards. As shown in Figure 1, Item 19 from Year 2003 has been tagged with "G.2.8 Congruence and similarity", the 2nd learning standard in the **G**eometry strand according to its congruence.

2.2 WPI Transfer Model

Several attempts have been made to use MCAS learning standards to code items, first with one standard per question, and then our own coding allowing each question to be tagged with multiple standards. Because we could not get statistically reliable coefficients on the learning standards, we hypothesize that a finer grained model would help. Additionally, more detailed analysis is needed for reporting to teachers and for predicting students' responses.

WPI300, containing only 174 knowledge components, is the first model created. Knowledge components are arranged in a hierarchy based on prerequisite structure. Questions point to an unordered list of knowledge components. To date, 102 knowledge components have been used to tag 92 assistments (including 853 questions). Generated reports reveal detailed information about students' learning and knowledge components contained in problems. Our current goal is to show that WPI300, as a finer grained cognitive model, will be more predictive. Math education researchers can upload their own transfer model for subsequent reporting.

3. REPORTING SYSTEM

3.1.1 Student Grade Book Report

Nine teachers using the Assistment system (every two weeks) since September, 2004value the system not only because their students get instructional assistance (e.g., scaffolding questions and hints), but also because they get online, live reports on students' progress.

The "Grade Book", shown in Figure 3.1, is the most frequently used report by teachers. Each row in the report represents information for one student, including total minutes using assistments, minutes today, number of problems and percent correct, prediction of MCAS score and performance level[2]. Grad Book also summarizes student actions in an "Assistment metric": number of scaffolding questions, student performance on scaffolding questions and number of student hint requests. In addition to performance, the "Assistment metric" exposes unusual student behaviour (e.g., making more attempts and requesting more hints than other students), suggesting students did not take assistments seriously or were "gaming the system" [1].

Student Name	Total time before (min)	Time spent today (min)	Original Items					Scaffolding + Orig. Items				Most Difficult MA. Standard
			# Done	# Correct	% Corr.	MCAS Score*	Perf. Level	# Done	# Correct	% Correct	# Hint Req.	
Tom	34	0	15	3	20%	200	Failing	30	16	53%	15	N.1.8-understanding-number-representations (Error times: 5/6)
Dick	32	0	38	26	68%	242	Proficient	81	56	69%	4	P.1.8-understanding-patterns (Error times: 2/6)
Harry	33	0	20	9	45%	220	Needs improv.	63	28	44%	63	P.1.8-understanding-patterns (Error times: 8/10)

FIGURE 3.1
Grade Book on real student data.

Figure 3.1 shows 3 students who used the system for about 30 minutes. (Many have used it for about 250 minutes). "Dick" finished 38 original items and only asked for 4 hints. Since most items were correct, his predicted MCAS score is high. He also made the greatest number of errors on questions tagged with the standard "P.1.8 understanding patterns". The student had done 6 problems tagged with "P.1.8" and made errors on 2 of those problems. "Harry" asked for numerous hints (63 compared to 4 and 15), so a teacher could confront the student with evidence of gaming or give him a pep-talk. By clicking a link to the student's name, teachers can see each action a student has made, the tutor's response and time spent on a given problem. The "Grade Book" is

[1]Our "prediction" of a student MCAS score is at this point primitive. The column is currently simply a function of percent correct. Given the nature of high stakes involved in MCAS, better predictions will be possible when students' real scores become available. We might even remove these two columns related to MCAS score prediction until we feel more confident in our prediction, in another word, "rough and ready".

so detailed that a student commented: “It’s spooky”, “He’s watching everything we do”.

The teacher can see individual questions associated with individual (e.g., difficult) knowledge components and the kind of errors made. (See Figure 3.2), thereby helping teachers to improve their instruction and to correct students’ misunderstandings.

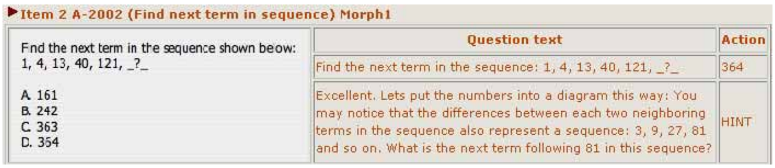


FIGURE 3.2
Items tagged with difficult knowledge component.

3.1.2 Class Summary Report

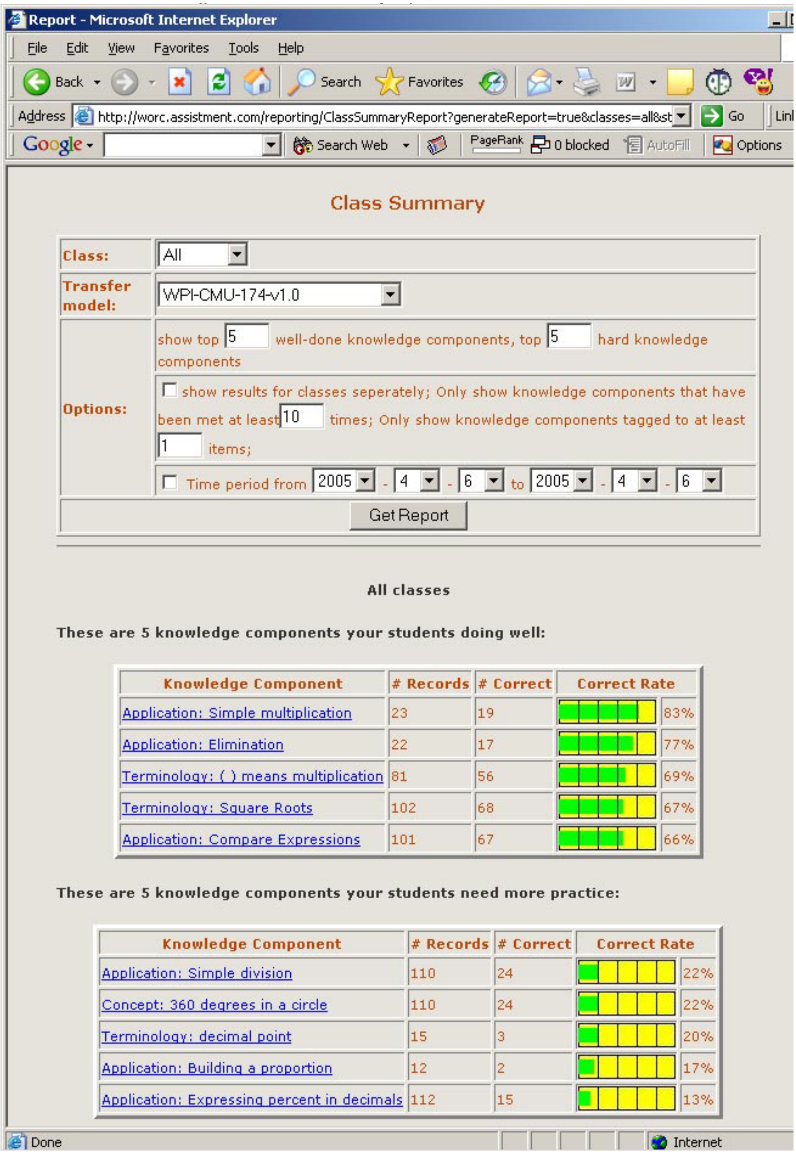
Class Summary” is a report informing teachers about the knowledge status of classes. Teachers select their favourite transfer models, specify how many and which knowledge components are to be shown in the report (e.g., by tagging).. They can also identify knowledge components on which students are good or bad during given time periods. Knowledge components are ranked according to correctness (green bars and percent correct in Figure 3.3,). Clicking the name of a knowledge component hyperlinkdirects teachers items tagged with the component. Teachers can see question and preview or analyze the item.

These reports help teachers decide which knowledge components and items to focus on to maximize students’ achievement.

3.1.3 Class Progress Report

Since teachers let their students only use the Assistment system every two or three weeks, we show teachers their students’ progress at days they worked on the assistments.

Figure 3.4 shows a preliminary progress reportfor a class using the system since September 21st, 2004 for a total of 9 times to date. Predicted average MCAS raw scores increased from 18 to 33, with 33 being relatively stable. [Note: These predicted MCAS scores are conservative — based on every items ever done, instead of only items done on lab days.]



Report - Microsoft Internet Explorer

http://worc.assistment.com/reporting/ClassSummaryReport?generateReport=true&classes=all&st

Class Summary

Class: All

Transfer model: WPI-CMU-174-v1.0

Options: show top 5 well-done knowledge components, top 5 hard knowledge components

show results for classes seperately; Only show knowledge components that have been met at least 10 times; Only show knowledge components tagged to at least 1 items;

Time period from 2005 - 4 - 6 to 2005 - 4 - 6

Get Report

All classes

These are 5 knowledge components your students doing well:

Knowledge Component	# Records	# Correct	Correct Rate
Application: Simple multiplication	23	19	83%
Application: Elimination	22	17	77%
Terminology: ( ) means multiplication	81	56	69%
Terminology: Square Roots	102	68	67%
Application: Compare Expressions	101	67	66%

These are 5 knowledge components your students need more practice:

Knowledge Component	# Records	# Correct	Correct Rate
Application: Simple division	110	24	22%
Concept: 360 degrees in a circle	110	24	22%
Terminology: decimal point	15	3	20%
Application: Building a proportion	12	2	17%
Application: Expressing percent in decimals	112	15	13%

FIGURE 3.3
Class summary report for a teacher's classes.

Class	Date	# Correct	# Total	# Student	Avg. Score	Std. Dev.	Confidence Interval(95%)
Period 3	2004-09-21	153	382	23	18	9.95	[-1.50 - 37.50]
Period 3	2004-10-27	427	773	23	25	11.18	[3.09 - 46.91]
Period 3	2004-11-10	630	1119	24	26	11.03	[4.37 - 47.63]
Period 3	2004-12-01	879	1437	22	29	10.20	[9.01 - 48.99]
Period 3	2004-12-15	1167	1790	21	32	8.24	[15.85 - 48.15]
Period 3	2005-02-02	1341	2029	20	33	7.96	[17.40 - 48.60]
Period 3	2005-02-16	1702	2576	23	33	6.67	[19.93 - 46.07]
Period 3	2005-03-02	1972	3065	24	33	6.61	[20.04 - 45.96]
Period 3	2005-03-16	2106	3288	23	33	6.58	[20.11 - 45.89]

FIGURE 3.4
Preliminary progress report for a class.

Item 20 N-2003 Morph (3/4 of 1 2/3)	24%
Item 20 N-2003 (2/3 of 1 1/2) Morph2	26%
Item 18 G-1998 (Angle in isosceles triangle)	27%
Item 35 G-2001 (Angle between clock hands)	27%
Item 13 D-1998 (Eiffel Tower model)	29%

FIGURE 3.6
Problems order by correct rate.

3.2 Analysis of Items

Item 20 N-2003 Morph (3/4 of 1 2/3) 24% Item 20 N-2003 (2/3 of 1 1/2) Morph2 26% Item 18 G-1998 (Angle in isosceles triangle) 27% Item 35 G-2001 (Angle between clock hands) 27% Item 13 D-1998 (Eiffel Tower model) 29% **Figure 3.6**: Problems order by correct rate

The report shows the average correct rate of each problem, enabling teachers easily compare problem difficulty (See Figure 3.6). By breaking original items into scaffolding questions and tagging scaffolding questions with knowledge components, one can analyze individual steps of a problem. Figure 3.4 is a scaffolding report showing statistics on each scaffolding question associated with any particular item.

---

[2]You may notice that 154 is less than 88% of 180, which should be about 158. And the number of attempts on later scaffolding questions went down more. That's because students could log out and log back in to redo the original question to avoid going through all scaffolding questions. We are trying to avoid this problem.

ID	Question	Correct Answer	% Correct	Hint Req.	# Attempt	Common Errors			WPI's Use of MA. Standard	WPI's Knowledge Components
						Resp.	#	Buggy Message		
Triangles ABC and DEF are congruent. The perimeter of triangle ABC is 23 inches. What is the length of side DF in triangle DEF?		10	12%	56%	180	8	15	N/A	**G.2.8**, M.3.8, P.7.8	Composition, T.3, A.3, T.4, A.4, A.12, A.15, A.17
						16	13	N/A		
						23	8	N/A		
1	Which side of triangle ABC has the same length as side DF of triangle DEF?	ac	23%	50%	154	ab	13	Side AB corresponds to side DE of triangle DEF, not DF. Try again, please.	**G.2.8-congruence-and-similarity**	Term: "Congruency", Appl: Congruency
						DF	6	N/A		
2	What is the perimeter of triangle ABC?	2x+x+8	39%	20%	148	2x + 8	69	No. It looks like you have added just two of the sides of triangle ABC. Perimeter is the sum of all the sides.	M.3.8-using-measurement-formulas	Term: "Perimeter", Appl: Perimeter
3	Now, given the perimeter of triangle ABC equals 23 inches, you can write the equation **2x + x + 8 = 23** and solve it for **x**. What is the value of **x**:	5	25%	52%	147	15	13	N/A	P.7.8-setting-up-and-solving-equations	Appl: Solve linear equation
						13	10	N/A		
						8	10	N/A		
4	Remember, we are looking for side DF. Enter the length of side DF:	10	30%	43%	143	5	26	N/A	G.2.8-congruence-and-similarity	Appl: Congruency
						2x	2	N/A		
						8	3	N/A		

FIGURE 3.7
A scaffolding report generated by Assistment reporting system.

The first line of Figure 3.7 shows a hard problem – which only 12% of students got correct on their first attempt. Only 154 of the 180 students[3] answered the original question correctly, forcing the system to introduce scaffolding questions. 56% of students asked for a hint, indicating low students' confidence. (Comparing such numbers across problems helps identify items on which students think they need help.) The state classified the item according to its "congruence" (G.2.8) shown in bold. Other MA learning standards (M.3.8, P.7.8) were added in our first attempt to code these MCAS 39 standards. Only 23% of students that got the original item incorrect can correctly answer the first scaffolding question — suggesting congruence is tough. The low percent correct (25%) on the 3rd question, asking students to solve for x, suggests varied sources of difficulty. Accordingly, we tagged "P.7.8-setting-up-and-solving-equations" to the problem.

Teachers want to know particular skills or knowledge components causing studentdifficulty in solving problems. Unfortunately the MCAS is not designed to be cognitively diagnostic and cannot help with such important questions. The scaffolding report provides less detailed cognitive diagnosis and cooperating teachers have designed scaffolding questions to help identify answers. For example, one teacher designed an assistment for ("What's ? of 1 ??"). The first scaffolding question for the assistment is "what mathematical operation does the word 'of' represent in the problem". This teacher said, "Want to see an item that 97% of my students got wrong?

Here it is… and it is because they don't know 'of' means they should multiply." Our reporting system confirmed the teacher's hypothesis. Over 40% of her students could not select "multiplication" with 11 selecting "division".

The scaffolding report has helped us develop our tutors in an iterative way. For each question, the report shows common errors and corresponding "buggy" messages. When building the Assistments, we have tried to "catch" common errors students might make and give them corrective feedback on that specific error. Because students may have different understandings of concepts, assistments may not give "buggy" messages for common errors, with our tutor loosing tutoring opportunities. Students may feel frustrated if they are continually being told "You are wrong" but get nothing instructive or encouraging. As shown in Figure 3.4, the wrong answer "15" to the third question has been given 13 times, but the assistment gave no feedback except correct or wrong. The assistment builders can improve their tutor by using online authoring tools [8] to add "buggy" messages like "It seems you have got the value of 2x+x, but we are looking for the value of x."

A table called "Red & Green" distribution matrix (Table 3.1) is also shown in the scaffolding report. Numbers in the cells show how many students got questions correct (indicted by green number in un-shaded cells) or wrong (indicated by red in shaded cells). The number is split as the questions' sequence number grows also showing how those students have done on previous questions. In this example, 4 students who answered the original question incorrectly went through all of the scaffolding questions correctly. This suggests that students have mastered the knowledge components required at each step and only need instruction on how to "compose" those steps. Eight students answered the original question wrong but answered the last (asking the same question) correctly. Since the assistment breaks the whole problem into scaffolding steps and gives hints and "buggy" messages, we would like to believe those students learned from working on the previous steps of this assistment.

3.3 Are Students Learning within the assistment system? Results

Some assistments are similar to others. They are either "morphs" of the original items or include the same knowledge components. Hence, we expect students working on assistments to do better on similar problems. Doing learning analysis by hand is both time consuming and fallible. So another aim of our reporting system is to automate the learning analysis process. Toward this end, the CONNECT/Java package provided by Insight S-PLUS was used to access the S-PLUS engine through Java, making it possible to send statistical analyzing commands from our web server to the S-PLUS engine, then receiving and presenting the results on JSP pages. Figure 3.8 shows the result generated based on May, 2004 data. The “fractionMult” Learning Opportunity Group (LOG), including three fraction multiplication problems, showed a significant increase on students’ percent correct from the first to the second opportunity. The two opportunities were compared in S-PLUS using an ANOVA (p = .039) and a t-test (p = .015). Similar analysis of 2005 data set with over 600 students and 30 LOGs show statistically significant student learning — at about 5% on their second opportunity [7].

3.4 Performance evaluation

Our reporting system was first used in May, 2004. It worked well and most reports at the class level could be generated in less than 10 seconds. It took 10 to 20 seconds to generate a scaffolding report at a “system” level. Performance went down when the number of recorded student actions increased past 1 million. “Grade Book” reports have taken more than 2 minutes, which is unacceptable as a live report. Switching to an Oracle database The approaches used to generate reports was also updated. Now, the “Grade Book” report can be generated in an average of about 7 seconds. The system level scaffolding report for Item 19 (See Figure 3.4) takes about 5 seconds.

4. CONCLUSIONS

In conclusion we are developing state-of-the-art online reporting tools to help teachers become better informed about what their students know. These reports appear to work live in the classroom. There is still much to be done in automating statistical analysis of learning experiments. Our long term vision is to let teachers create content, and to let them know automatically by email when it has been determined that their content is better (or worse) than what is used currently. We have taken initial stops in that direction.

REFERENCES

[1] Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004) Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, 2004, Maceio, Brazil

[2] Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

[3] Confrey, J., Valenzuela, A. & Ortiz, A. (2002). Recommendations to the Texas State Board of Education on the Setting of the TAKS Standards: A Call to Responsible Action. At http://www.syrce.org/State_Board.htm

[4] Croteau, E., Heffernan, N. T. & Koedinger, K. R. (2004) Why Are Algebra Word Problems Difficult? Using Tutorial Log Files and the Power Law of Learning to Select the Best Fitting Cognitive Model, *Proceedings of the 7th International Conference on Intelligent Tutoring System*, 2004, Maceio, Brazil

[5] Nuzzo-Jones, G., Walonoski, J.A., Heffernan, N.T., Livak, T. (2005). The eXtensible Tutor Architecture: A New Foundation for ITS. *Submitted to the 12th Annual Conference on Artificial Intelligence in Education* 2005, Amsterdam

[6] J. Mostow, J.E. Beck, R. Chalasani, A. Cuneo, and P. Jia. Viewing and Analyzing Multimodal Human-computer Tutorial Dialogue: A Database Approach. *Fourth IEEE International Conference on Multimodal Interfaces* (ICMI 2002), October, 2002.

[7] Razzaq, L, Feng, M., Nuzzo-Jones, G., Heffernan, N.T., Aniszczyk, C., Choksey, S., Livak, T., Mercado, E., Turner, T., Upalekar, R., Walonoski, J., Macasek, M., Rasmussen, K., Koedinger, K., Junker, B., Knight, A., Ritter, S. (2005). The Assistment Project: Blending Assessment and Assisting. *Submitted to the 12th Annual Conference on Artificial Intelligence in Education* 2005, Amsterdam

[8] Turner, T., Macasek, M.A., Nuzzo-Jones, G., Heffernan, N.T. (2005). The Assistment Builder: A Rapid Develoment Tool for ITS. *Submitted to the 12th Annual Conference on Artificial Intelligence in Education* 2005, Amsterdam