Kong, S.C., Ogata, H., Arnseth, H.C., Chan, C.K.K., Hirashima, T., Klett, F., Lee, J.H.M., Liu, C.C., Looi, C.K., Milrad, M., Mitrovic, A., Nakabayashi, K., Wong, S.L., Yang, S.J.H. (eds.) (2009). *Proceedings of the 17th International Conference on Computers in Education [CDROM]*. Hong Kong: Asia-Pacific Society for Computers in Education.

Team Neurodynamics
by The
Learning Chameleon, Inc

Tracking the Development of Problem Solving Skills with Learning Trajectories

Ron STEVENS[a], Carole BEAL[b], Marcia SPRANG[c]
[a]*UCLA IMMEX Project, 5601 W. Slauson Ave. #255, Culver City, CA 90230 USA.*
[b]*College of Social & Behavioral Sciences, University of Arizona, P.O. Box 210077 Tucson, AZ 85721 USA.*
[c]*Placentia-Yorba Linda Unified School District, 1830 N. Kellogg Drive Anaheim, CA 92807 USA.*
[a]immex_ron@hotmail.com

Abstract: Learning trajectories have been developed for 1650 students who solved a series of online chemistry problem solving simulations using quantitative measures of the efficiency and the effectiveness of their problem solving approaches. These analyses showed that the poorer problem solvers, as determined by item response theory analysis, were modifying their strategic efficiency as rapidly as the better students, but did not converge on effective outcomes. This trend was also observed at the classroom level with the more successful classes simultaneously improving both their problem solving efficiency and effectiveness. A strong teacher effect was observed, with multiple classes of the same teacher showing consistently high or low problem solving performance.
The analytic approach was then used to better understand how interventions designed to improve problem solving exerted their effects. Placing students in collaborative groups increased both the efficiency and effectiveness of the problem solving process, while providing pedagogical text messages increased problem solving effectiveness, but at the expense of problem solving efficiency.

Keywords: Problem solving; learning trajectories; pedagogical agents; artificial neural networks.

Introduction

Technology-based learning environments provide the foundation for a new era of integrated, learning-centered assessment systems [1]. It is now becoming possible to rapidly acquire data about students' changing knowledge, skill and understanding as they engage in real-world complex problem solving, and to create predictive models of their performance both within problems [2] as well as across problems and domains [3]. A range of analytic tools are being applied in these analyses including Bayesian Nets [4], computer adaptive testing based on item response theory (IRT) [5], regression models and artificial neural networks (ANN) [6], [7], each of which possesses particular strengths and limitations [8].
How can this data be best put to use? Recent analyses of traditional assessment approaches and professional development models indicate that interventions often fail because teachers either do not fully understand how to implement them, or are not adequately supported in their efforts to implement them [9], [10], [11]. Simply increasing teachers' access to assessment data however, may only exacerbate the challenges that they face in crowded classrooms when adapting instruction. Thus, new approaches are needed to

provide teachers with accurate, predictive and useful data about their students' learning in ways that are easily and rapidly understood. Data available in real time, that speak to process as well as outcomes, and that are intuitively easy to understand would seem to be minimum requirements.

Finding the optimum granular and temporal resolutions for reporting this assessment data will be a fundamental challenge for making the data accessible, understandable and useful for a diverse audience (e.g. teachers, policy makers and students) as each may have different needs across these dimensions [12], [13]. If the model resolution is general and / or delayed then important dynamics of learning may be lost or disguised for teachers. If the resolution is too complex or the reporting too frequent the analysis will become intrusive and cumbersome.

Methods

We have been developing reporting systems for problem solving which are helping to measure how strategically students are thinking about scientific problems and whether interventions to improve this learning are having the desired effect. The system is termed IMMEX (Interactive MultiMedia Exercises), an online library of problem solving science simulations that is coupled with layers of probabilistic tools which assess students' problem solving performance, progress, and retention [7],[14], [15], [16], [17], [18]. One IMMEX task is called *Hazmat*, which provides evidence of a student's ability to conduct qualitative chemical analyses [19]. IMMEX problems are what Frederiksen [20] referred to as "structured problems requiring productive thinking", meaning that the problems can be solved through multiple approaches, and students cannot rely on known algorithms to decide which resources are relevant and how the resources should be used.

Hazmat contains 38 problem cases which involve the same basic scenario but vary in difficulty due to the properties of the different unknown compounds being studied. These multiple instances provide many opportunities for students to practice their problem solving and also provide data for Item Response Theory (IRT) estimates of problem solving ability which can be useful for comparing outcomes with more traditional ability measures such as grades.

IMMEX supports detailed assessments of students' overall problem solving effectiveness and efficiency by combining solution frequencies (or IRT estimates) which are outcome measures and artificial neural network (ANN) and hidden Markov modeling (HMM) performance classifications which provide a strategic dimension [21], [22], [23], [24]. To simplify reporting and to make the models more accessible for teachers, these different layers of data can be combined into an economics-derived approach which considers students' problem solving decisions in terms of the resources available (what information can be gained) and the costs of obtaining the information. Students who review all available problem resources are not being very efficient, although they might eventually find enough information to arrive at the right answer. Other students might not look at enough resources to find the information required to solve the problem, i.e., they are being efficient but at the cost of being ineffective. Students demonstrating high strategic efficiency should make the most effective problem-solving decisions using the fewest number of the resources available. As problem solving skills are gained this should be reflected as a process of resource reduction (i.e. higher efficiency) and improved outcomes (greater effectiveness) [25]. Over arching these dimensions of efficiency and effectiveness is a student's content knowledge measured by conventional testing practices.

Kong, S.C., Ogata, H., Arnseth, H.C., Chan, C.K.K., Hirashima, T., Klett, F., Lee, J.H.M., Liu, C.C., Looi, C.K., Milrad, M., Mitrovic, A., Nakabayashi, K., Wong, S.L., Yang, S.J.H. (eds.) (2009). *Proceedings of the 17th International Conference on Computers in Education [CDROM]*. Hong Kong: Asia-Pacific Society for Computers in Education.

Results

The data gathered as students work with IMMEX provide rich, real-time assessment information along the efficiency and effectiveness dimensions. Figure 1 shows a modeling across schools and teachers/classrooms (66 classrooms, 62,774 performances) where an index of strategic efficiency [26] is plotted against an effectiveness (i.e. solution frequency) rate. The quadrants generated by intersections of the averages of these measures reflect 1) mostly guessing (upper left corner), 2) performances where students view many resources, but miss the solution (lower left), 3) performances where many resources are being viewed and the problem is being solved (lower right) and 4) the performances where few resources are used and the problem is solved (upper right). As expected by the visualization format, schools are distributed across the quadrants (Figure 1, left). A second level of analysis showing problem solving performance across five teachers as well as their classrooms where the different classes of the same teacher are shown by the symbols, and the different teachers identified by numbers (Figure 1, right). The clustering of the different classrooms of the teachers (for instance, the +'s in the lover left hand corner and the squares in the upper right corner), illustrates a significant teacher effect perhaps reflecting different instructional methods [25]. Followup classroom observation studies by Thadani et al [18] suggest that the teacher's mental model of the problem space, and approach for solving the problem, can have a major effect on the approach adopted by the students.

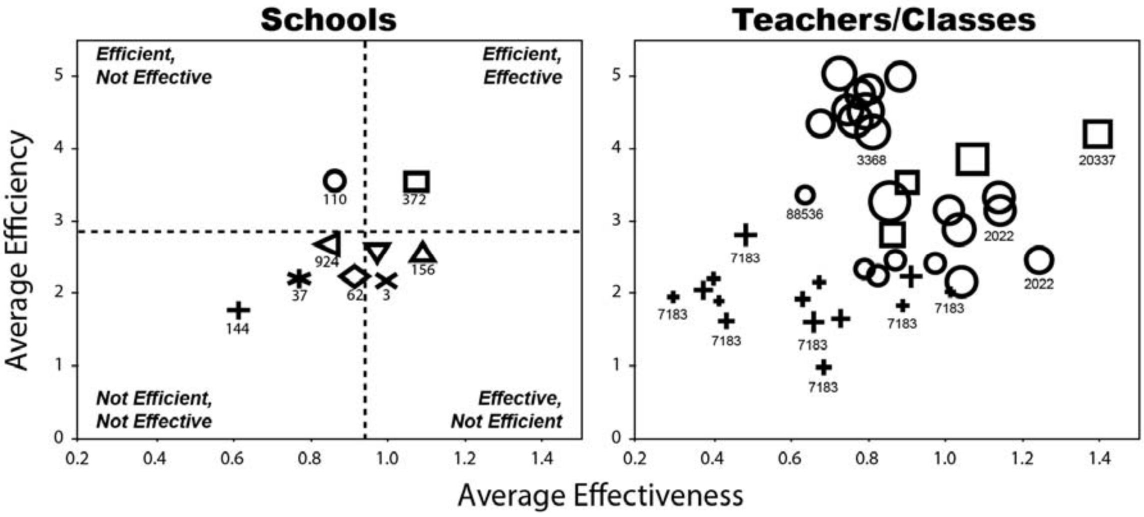

Figure 1. Aggregated Efficiency and Effectiveness measures of Schools and Classrooms that Performed *Hazmat*. The dataset was aggregated by schools (*left*) and then by teachers (symbols and text) and classrooms (*right*) and the efficiency (on a scale of 0-6) and effectiveness (on a scale of 0-2) measures calculated as described previously [19], [20]. The symbol sizes are proportional to the number of performances. Each axis in Figure 1A is bisected by dotted lines indicating the average efficiency and effectiveness measures of the dataset creating quadrant combinations of high and low efficiency and effectiveness.

Tracking problem solving efficiency and effectiveness as multiple *Hazmat* problems are performed creates a learning trajectory (Figure 2) which is an important formative assessment tool showing how students improve with practice [27].

Learning trajectories show that the poorer problem solvers, as determined by IRT analysis, are modifying their strategic efficiency as rapidly as the better students, as shown by the position changes along the Efficiency axis, but they are not converging on effective outcomes (Figure 2A). Figure 2B shows that this trend can be observed in classrooms as well, (e.g. Class 1). While the more successful classes (e.g. Class 4) simultaneously

Kong, S.C., Ogata, H., Arnseth, H.C., Chan, C.K.K., Hirashima, T., Klett, F., Lee, J.H.M., Liu, C.C., Looi, C.K., Milrad, M., Mitrovic, A., Nakabayashi, K., Wong, S.L., Yang, S.J.H. (eds.) (2009). *Proceedings of the 17th International Conference on Computers in Education [CDROM]*. Hong Kong: Asia-Pacific Society for Computers in Education.

improved both their problem solving efficiency and effectiveness, the lower performing classes showed gains only in efficiency The learning trajectories are also important as changes in problem solving progress can be detected after as few as 2-4 student performances providing an opportunity for intervention before poor approaches have been learned. For instance, a teacher could initiate an intervention with a smaller group of students and after they have performed part of their assignment the teacher can observe online whether this was having a positive, negative or neutral effect and either continue or modify the intervention.

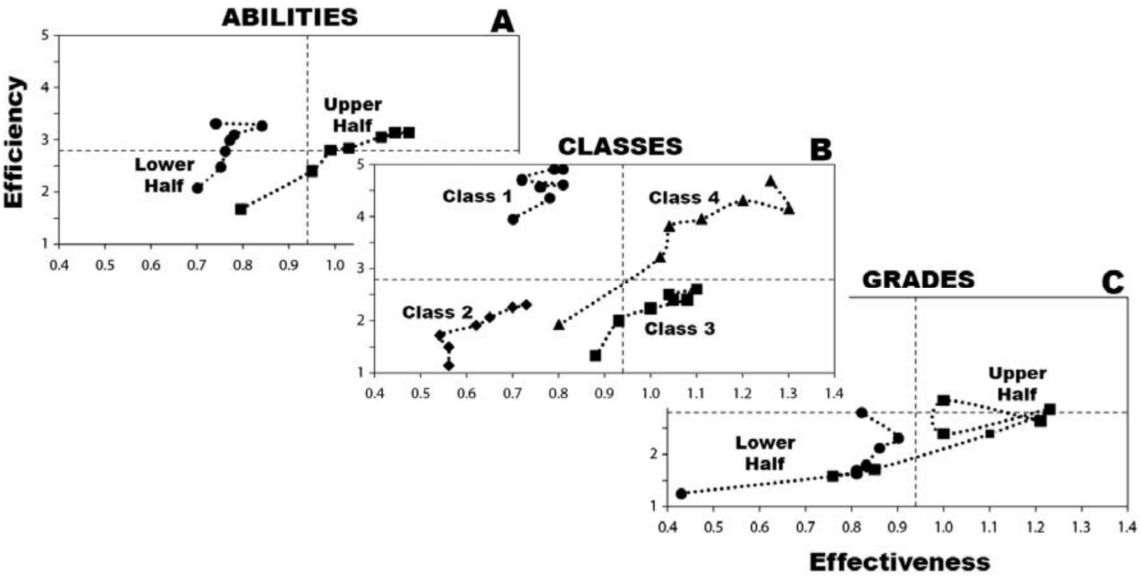

Figure 2. Learning Trajectories of Classes and Students of Different Abilities. A) The dataset (n = 62,774) was divided into lower (IRT scores = 3.4 to 49.3) and higher (IRT scores = 49.4 to 60.3) Hazmat problem solving ability students and the learning trajectories plotted. B) The Efficiency / Effectiveness measures are stepwise plotted for 7 Hazmat performances for four representative classes. C) A dataset (82 students, 780 Hazmat performances) for three Advanced Placement Chemistry classes was divided into high and low categories based on the final course grade and the learning trajectories calculated.

A similar analysis was conducted for 80 students in three Advanced Placement Chemistry students who were separated into the upper and lower halves based on their final course grades. Again, the learning trajectories of the lower half of the students showed similar increases in strategic efficiency as the upper half of the students, but remained lower in effectiveness [28]. Thus from the perspectives of problem solving abilities, course grades, and perhaps the instructional environment it would appear that some students may have difficulty monitoring their problem solving resulting in decreased outcomes and that interventions designed to improve monitoring skills may be useful.

From a formative assessment perspective learning trajectories can provide evidence as to whether interventions adopted to improve learning are working. The learning trajectory for students (N = 50,062 performances, -○-) who improved at their own pace is characterized by progressive improvement across both the efficiency and effectiveness dimensions which begins to plateau after around 4 performances (Figure 3). This plateau mirrors the stabilization of strategies and abilities we have previously documented using HMM and IRT [21], [22], [23].

A second learning trajectory is from students who received text messages that were integrated into the prologue of each problem, i.e., before the student began actually working on the problem, that were designed to encourage students to reflect on their problem solving (n = 11,497 performances, -□-). The messages suggested, for example:

"When you read the IMMEX problem, don't let yourself rush into trying different things. Stop and think for a minute first. What have you learned in science class that could help you identify the right place to start?" Students who received the metacognitive hints became less efficient, meaning that they looked at more problem materials, but they also became more effective problem solvers, i.e. they become more cautious or reflective. A control group of students (n = 1,215 performances, -●-) also received messages, but here the messages were designed to be generic academic advice (e.g., "It's a good idea to keep up with the reading for your science class."). These students became less efficient as well as less effective, i.e. these messages may have been a distraction from their problem solving. Thus, the message content was critical to improving students' problem solving; the presence of text messages alone was not helpful. Finally, grouping students into pairs (n = 5,577 performances, -■-), improved both the efficiency as well as the effectiveness of the problem solving strategies.

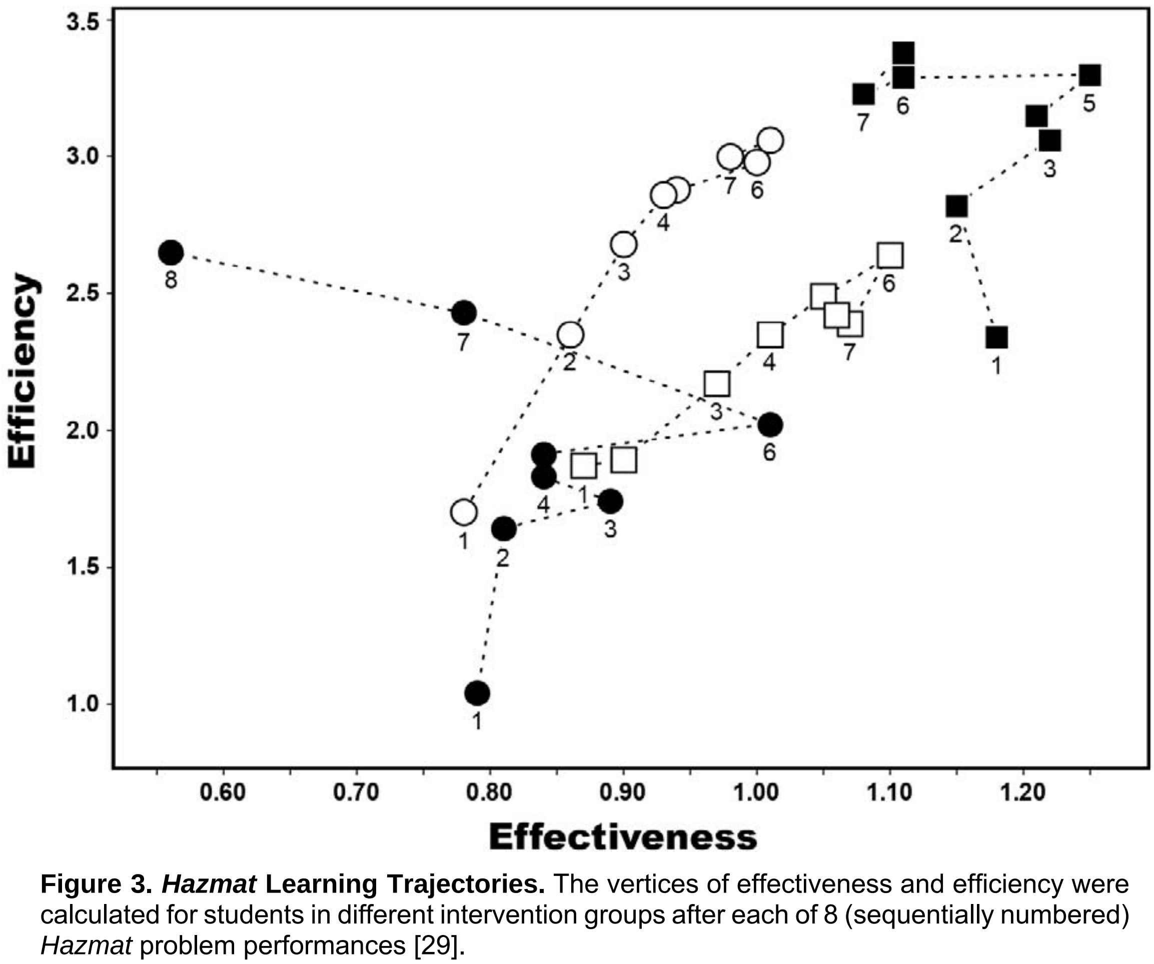

Figure 3. *Hazmat* Learning Trajectories. The vertices of effectiveness and efficiency were calculated for students in different intervention groups after each of 8 (sequentially numbered) *Hazmat* problem performances [29].

We recognize that the ultimate power of improved measures of student learning cannot be realized if the data presentation for the teacher (and student) does not provide both compelling and readily understood results. The online IMMEX Digital Dashboard that is now available online is shown in Figure 4. The idea is that teachers gain access to the online data mining control as part of their normal login process. The starting point for the digital dashboard interface presents a rose petal diagram where each leaf represents a class and the length of each petal indicates the number of problem solving performances in each class; the shaded area shows the percentage of cases solved. In Figure 4 for example, this teacher has 4 classes, and by comparing the shaded areas in the classes represented at 6 o'clock and 9 o'clock it is apparent that the first class is solving more cases than the second.

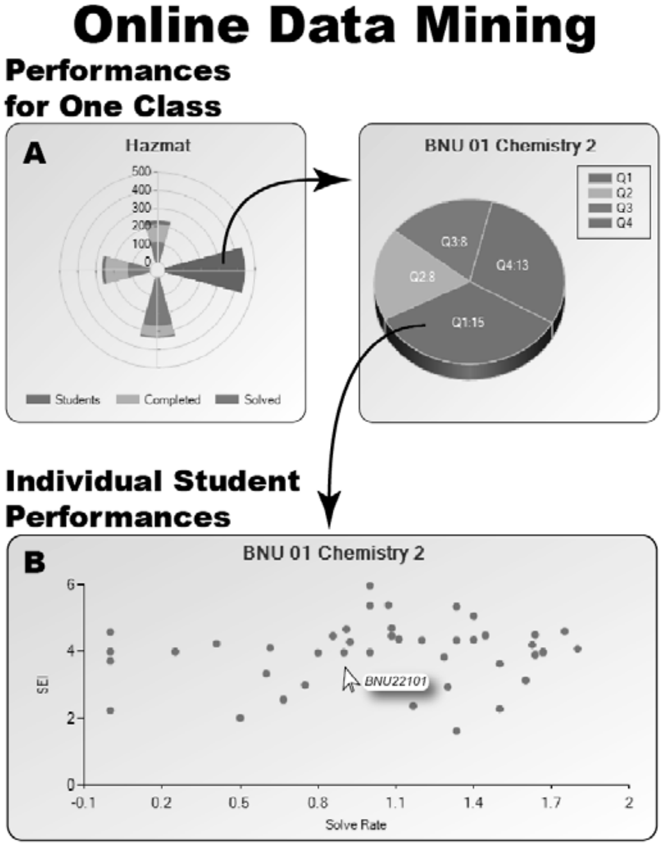

Figure 4. Sample Layout for IMMEX Data Dashboard for Reporting Student Performance to Teachers

Clicking on each petal (shown for the class at 3 o'clock) brings up a display for that class showing the distribution (described in Figure 1) of students across the four efficiency / effectiveness quadrants which when clicked again provides details for each student in the class. Teacher's attention is therefore drawn to students requiring specific forms of help and this can be supplemented by incorporating intelligent annotations highlighting subsections of data. After viewing this information, the teacher may choose to provide a form of differentiated instruction to individual students, groups of students or entire classes before having them continue to problem solve.

Discussion

These studies show that technology can provide dynamic models of what students are doing as they learn problem solving without creating a burden on educational systems. While illustrated for chemistry, such models are applicable to other problem solving systems where learning progress is tracked longitudinally. When shared with teachers and students in real time [30] they can provide a roadmap for better instruction by highlighting problem solving processes and progress and documenting the effects of classroom interventions and instructional modifications. The differences observed across schools, teachers and student abilities shifts the focus to the classroom and may provide a means for matching students and instruction or matching teachers with professional development activities.

Acknowledgements

Supported in part by NSF grants ROLE 0528840 and DUE 0512526

References and Footnotes

[1] Quellmalz, E. S., Pellegrino, J. W. (2009). Technology and Testing. *Science* Vol. 323. no. 5910, pp. 75-79.

[2] Murray, R. C. & VanLehn, K. (2000). A Decision-Theoretic, Dynamic Approach for Optimal Selection of Tutorial Actions. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent Tutoring Systems*, Fifth International Conference, ITS 2000, pp. 153-162. Montreal, Canada. New York: Springer.

[3] Stevens, R. H., Soller, A., Cooper, M., and Sprang, M. (2004). Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor. *Intelligent Tutoring Systems.* J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds). Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings (pp. 580-591).

[4] Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K. B. Laskey & H. Prade (Eds.), *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (pp 437-446). San Francisco: Morgan Kaufmann.

[5] Linacre, J.M. (2004). WINSTEPS Rasch measurement computer program. Chicago. Winsteps.com

[6] Beal, C. R., Mitra, S., & Cohen, P. R. (2007). Modeling learning patterns of students with a tutoring system using Hidden Markov Models. Proceedings of the13th International Conference on Artificial Intelligence in Education, IOS Press.

[7] Soller, A., and Stevens, R. H. (2007). Applications of Stochastic Analyses for Collaborative Learning and Cognitive Assessment. In Advances in Latent Variable Mixture Models, Gregory Hancock and Karen Samuelson (Eds.). Information Age Publishing

[8] Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (Eds.) (2006). Automated scoring of complex tasks in computer based testing. Mahwah, NJ: Erlbaum Associates.

[9] Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433-479.

[10] Lawless, K. A. & Pellegrino, J. W. (2007). Professional Development in Integrating Technology Into Teaching and Learning: Knowns, Unknowns, and Ways to Pursue Better Questions and Answers. *AERA Review of Educational Research*, Vol. 77, No. 4, 575-614.

[11] Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy Implementation and Cognition: Reframing and Refocusing Implementation Research. *Review of Educational Research,* 72 (3), pp. 387-431.

[12] Alberts, B. (2009). Redefining Science Education. *Science*, Vol. 323. no. 5913, p. 437.

[13] Loehle, C. (1990). A Guide to Increased Creativity in Research – Inspiration or Perspiration. Bioscience Vol. 40, no. 2, pp: 123-129.

[14] Stevens, R. H, & Palacio-Cayetano, J. (2003). Design and Performance Frameworks for Constructing Problem-solving Simulations. *Cell Biology Education*, 2 (3), pp. 162-179.

[15] Stevens, R. H., Soller, A., Cooper, M., and Sprang, M. (2004) Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor. *Intelligent Tutoring Systems.* J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds). Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings (pp. 580-591).

[16] Stevens, R. H., Wang, P., & Lopo, A. (1996). Artificial Neural Networks Can Distinguish Novice and Expert Strategies During Complex Problem Solving. *JAMIA*, Vol. 3, 2 pp 131-138.

[17] Cooper, M. M. Cox Jr., C. T., Nammouz, M., Case, E., & Stevens, R. H. (2008). An Assessment of the Effect of Collaborative Groups on Students' Problem Solving Strategies and Abilities. J. Chem Ed Vol. 85, No. 6, pp 866-872.

[18] Thadani, V., Stevens, R. H., & Tao, A. (2009 in press). Measuring Complex Features of Science Instruction: Developing Tools To Investigate the Link Between Teaching and Learning. *Journal of the Learning Sciences.*

[19] The task is described in detail, and the normalized performance dataset can be found at http://www.immex.ucla.edu/psom

[20] Frederiksen N. (1984). Implications of Cognitive Theory for Instruction in Problem Solving. *Review of Educational Research*, Vol. 54, No. 3, pp. 363-407 Published by: American Educational Research Association

[21] Stevens, R. H. (2007). A Value-Based Approach for Quantifying Student's Scientific Problem Solving Efficiency and Effectiveness Within and Across Educational Systems. In *Assessing and Modeling Cognitive Development in School.* (Robert W. Lissitz, ed). JAM Press, Maple Grove, MN. Pp217-240.

[22] Stevens, R. H., & Thadani, V. (2007). Quantifying Student's Scientific Problem Solving Efficiency and Effectiveness. *Technology, Instruction, Cognition and Learning (TICL)* Volume 5, Number 2-3-4, pp 325-337.

[23] Stevens, R. H., & Casillas, A. (2006). Artificial Neural Networks. In R. E. Mislevy, D. M. Williamson, & I. Bejar (eds), Automated Scoring of Complex Tasks in Computer Based Testing: An Introduction. Lawrence Erlbaum, Mahwah, NJ. (pp. 259-312).

[24] Haider, H., and Frensch, P.A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology* 30 (3): 304-337

[25] The core components of strategic efficiency are 1) the quantity of resources used vs. the quantity available, 2) the value of the resulting outcomes expressed as a proportion of the maximum outcomes, and 3) the quality of the data accessed.

[26] Zimmerman, C. (2007). The Development of Scientific Thinking Skills in Elementary and Middle School. Developmental Review. Vol. 27, no. 2, pp 172-223.

[27] Lajoie, S. P. (2003). Transitions and Trajectories for Studies of Expertise. *Educational Researcher.* 32 (8), pp. 21–25.

[28] The correlations between the final grades and the efficiency index, ability estimates by IRT, and the solved rates (i.e. effectiveness) were $R2 = 0.06$, $p = 0.02$, $R2 = 0.006$, $p = 0.49$, $R2 = 0.02$, $p = 0.23$.

[29] The Kruskal-Wallis One Way Analysis of Variance on Ranks showed differences in the median values among the treatment groups that are greater than would be expected by chance ($P = <0.001$) for both Efficiency and Effectiveness. All pairwise multiple comparisons (Dunn's Method) were significantly different ($P <0.05$).

[30] For information on real-time reporting, see: http://www.immex.ucla.edu/psom/dashboard_instructions.pdf (accessed 3/24/08).