A Bayesian Network Approach for Modeling the Influence of Contextual Variables on Scientific Problem Solving

Ronald H. Stevens[1], Vandana Thadani[2]
[1] UCLA IMMEX Project, 5601 W. Slauson Ave. #255, Culver City, CA 90230
immex_ron@hotmail.com
[2] Loyola Marymount University, Dept of Psychology, Los Angeles, CA 90045
vthadani@lmu.edu

Abstract. This paper describes the causal relationships between students' problem-solving effectiveness (i.e. reaching a correct solution) and strategy (i.e. approach) and multiple contextual variables including experience, gender, classroom environment, and task difficulty. Performances of the IMMEX problem set *Hazmat* (n~33,000) were first modeled by Item Response Theory analysis to provide a measure of effectiveness and then by self-organizing artificial neural networks and hidden Markov modeling to provide measures of strategic efficiency. Correlation findings were then used to link the variables into a Bayesian network representation. Sensitivity analysis indicated that whether a problem was solved or not was most likely influenced by findings related to the problem under investigation and the classroom environment while strategic approaches were most influenced by the actions taken, the classroom environment and the number of problems previously performed. Subsequent testing with unknown performances indicated that the strategic approaches were most easily predicted (17% error rate), whereas whether the problem was solved was more difficult (32% error rate).

1 Introduction

Strategic problem solving is a complex process with skill development being influenced by the task, the experience and knowledge of the student, the balance of cognitive and metacognitive skills possessed by the student, gender [2], ethnicity, classroom environment [10] and ability constructs such as motivation and self efficacy [8]. The variable contributions of these influences helps account for why it is so challenging for teachers to identify which students are using the knowledge and critical thinking skills presented in class to solve real-world problems, and distinguish them from other students that may require interventional supports [6]. These analyses are further complicated as the acquisition of problem solving skills is a dynamic and often gradual process characterized by transitional changes over time as experience is gained and learning occurs [4]. Given the nature of novice learning, student trajectories are likely to be complex with regard to the heterogeneity of strategies, the pace of learning, and the level of expertise obtained. [1].

To address these challenges we have been developing probabilistic models of learning trajectories that position students' scientific problem-solving skills upon a continuum of experience. These models provide estimates of student ability (Item Response Theory (IRT) analysis), describe the strategy used during problem solving session (Artificial Neural Networks (ANN)) and define trajectories of progress as multiple problems are performed (Hidden Markov Modeling (HMM)) [16].

A consistent finding across the domains of chemistry, molecular genetics, genetics, medicine and K-12 science is that after a period of practice students stabilize with a level problem solving competency characterized by particular approaches [13] [15] [16] [19] [21]. Furthermore, once stabilization has occurred, many students will use these approaches when presented with similar problems up to 3 months later. Unfortunately, not all students will stabilize with efficient and/or effective approaches indicating that experience alone is not sufficient for some students to progress, a finding reported by others [7]. The challenge therefore, is to rapidly identify students who are unlikely to make progress on their own and then target deliberate practice [1], teacher guidance, and/or interventions such as pedagogical feedback or collaborative group learning to improve the level of competency.

To enable predictive modeling it will be important to better understand how the diverse set of individual and contextual variables associated with complex problem solving differentially contribute to the adoption and persistence of strategies. In this paper we describe the construction and preliminary validation of descriptive Bayesian networks that can serve both as an analytic workbench to better understand the interactions among these variables, as well as an engine for developing support decisions for future problem solving and learning activities.

2 Methods

IMMEX (Interactive Multi-Media Exercises) is an online problem solving environment and layered analytic system that delivers problem solving tasks that require students to analyze descriptive scenarios, judge what information is relevant, plan a search strategy, gather information, and eventually reach a decision(s) that demonstrates understanding [20].

Since online delivery of these cases began 5 years ago, over 500,000 problems have been performed by students spanning middle school to medical school. One, of several problem sets researched extensively is *Hazmat*, which provides evidence of students' ability to conduct qualitative chemical analyses [16]. A multimedia presentation is shown to the students, explaining that an earthquake caused a chemical spill in the stockroom and their task is to identify the unknown chemical by gathering information using a 22 item menu containing a Library of terms, a Stockroom Inventory, and different Physical or Chemical Tests. This problem set contains 38 cases that can be performed in class, assigned as homework, or used as quizzes.

To follow students' performance and progress we have developed analytic models of how strategies are constructed, modified and retained as students learn to solve problems like *Hazmat* [16].

2.1 Model 1. Item Response Theory (IRT) Estimates of Student Ability.

The 38 Hazmat cases include a variety of acids, bases, and compounds giving either a positive or negative result when flame tested. As expected, the flame test negative compounds are more difficult for students because both the anion and cation need to be identified by running additional chemical tests. As students perform multiple cases, estimates of their ability can be obtained by IRT analysis that relates characteristics of items and individuals to the probability of solving a given case [5].

Overall, the problem set presents an appropriate range of difficulties to provide reliable estimates of student ability [19]. In the subsequent BN models we refer to these values as **IRT**.

2.2 Model 2. Artificial Neural Network (ANN) Classification of Strategies.

While useful for ranking the students by the outcomes of their problem solving, IRT does not provide strategic measures. Here, we use ANN analysis. As students navigate the problem spaces, the IMMEX database collects timestamps of each student selection. The most common student approaches (i.e. strategies) for solving *Hazmat* are identified with competitive, self-organizing artificial neural networks [3] [18] [15] using these time stamped actions as the input data. The result is a topological ordering of the neural network nodes according to the structure of the data where geometric distance becomes a metaphor for strategic similarity. Often we use a 36-node neural network and the details are visualized by histograms showing the frequency of items selected for student performances classified at each node (Figure 1a). **Strategies** so defined consist of actions that are always selected for performances at that node (i.e. with a frequency of 1) as well as ones ordered variably.

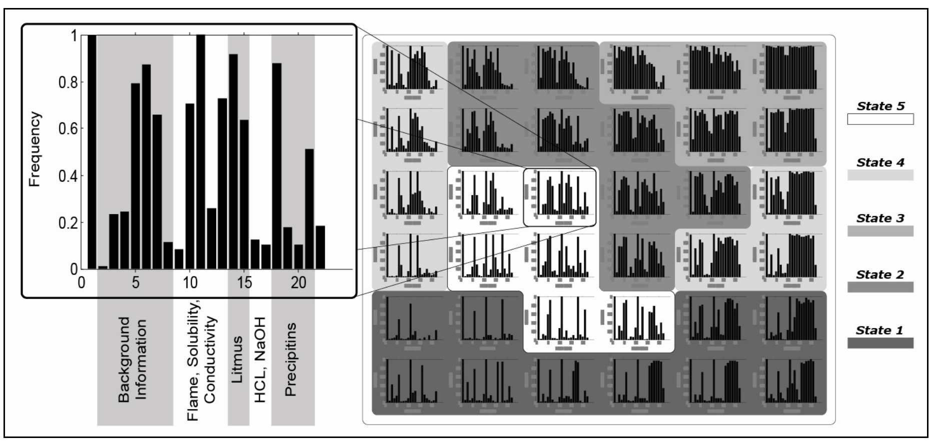

Figure 1. Sample Neural Network Nodal Analysis. a.) The selection frequency of each action (identified by the labels) is plotted for the performances at node 15, and helps characterize the performances clustered at this node and for relating them to performances at neighboring nodes. The nodes are numbered in rows, 1-6, 7-12, etc. **b.)** This figure shows the item selection frequencies for all 36 nodes.

Figure 1b is a composite ANN nodal map that shows the topology of performances generated during the self-organizing training process. Each of the 36 matrix graphs represents one ANN node where similar student's problem solving performances have become competitively clustered, and as the neural network was trained with vectors representing student actions, it is not surprising that a topology developed based on the quantity of items. For instance, the upper right of the map (nodes 6, 12) represents strategies where a large number of tests were ordered, whereas the lower left contains strategies where few tests were ordered.

2.3 Model 3. Hidden Markov Model (HMM) Strategic Progress Models.

On their own, artificial neural network analyses provide point-in-time snapshots of students' problem solving. More complete models of student learning should also account for the changes of student's strategies with practice. Here we postulate that students will pass through a number (3-5) of **States** as they shift their problem solving strategies over time. In these models students perform multiple cases in the 38-case Hazmat problem set, and each performance is classified with the trained ANN. Predictive models of student progress are then developed from sequences of these strategies with HMM [11] [9]. This results in a Transition Matrix, and an Observation Matrix representing the resulting model. This approach is shown in Figure 2 where students solved 6 Hazmat cases. One level (stacked bar charts) shows the distribution of the 5 HMM states across the 6 performances. On the first case, when students are framing the problem space, the two most frequent states are States 1 and 3. Moving up an analytical layer from HMM states to ANN nodal strategies (the 6 x 6 histogram matrices) shows that State 3 represents strategies where students ordered all tests, and State 1 where there was limited test selection. With experience the students transited from State 3 (and to some extent State 1), through State 2 and into States 4 and 5, the more effective states. By the fifth performance the State distributions stabilized after which time students without intervention tended not to switch their strategies, even when they were ineffective. Stabilization with ineffective strategies is of concern as students tend to retain their adopted strategies over at least a 3-months period [18].

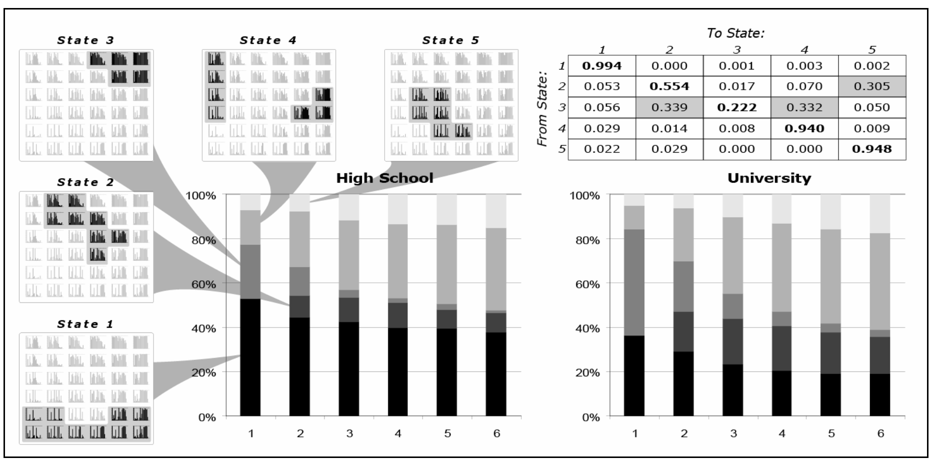

Figure 2. Modeling Individual and Group Learning Trajectories. This figure illustrates the strategic changes as high school and university students gain experience in *Hazmat* problem solving. Each stacked bar shows the distribution of HMM states for the students (N=7290) after a series (1-6) of performances. These states are also mapped back to the 6 x 6 matrices which represent 36 different strategy groups identified by self organizing ANN. The highlighted boxes in each neural network map indicate which strategies are most frequently associated with each state. From the values showing high cyclic probabilities along the diagonal of the HMM transition matrix (upper right), States 1, 4, and 5 appear stable, suggesting once adopted, they are continually used. In contrast, students adopting State 2 and 3 strategies are more likely to adopt other strategies (gray boxes).

3 Results

Crosstabulation analyses of over 75,000 student performances across multiple domains have repeatedly shown significant associations among student and contextual variables that influence both the problem-solving performance (solve rate, IRT ability estimates) as well as the approaches (ANN and HMM classifications) students adopt [12] [15] [17]. These variables include gender, the number of prior cases performed, the experience of the student (regular high school, AP high school, university), teacher and classroom effects as well as the conditions under which problem solving is performed (individual vs. collaborative).

Using commercial Bayesian network (BN) software (Netica, Inc), we have captured these interactions into belief networks to better understand the dependencies of the different variables. A sample BN is shown in Figure 3 where the network was initialized with a model of student problem solving based on a dataset of >33,000 *Hazmat* performances. Cross tabulation analysis has shown that dependencies exist among multiple nominal variables related to problem solving. The variables can be divided into two major categories: 1) dependent outcome measures that consist of whether or not the problem was solved (Solved/Not Solved in Figure 3), along with how the problems were solved (Strategy, State), and 2) contextual variables that include gender, the problem cases, student experience, the learning environment (individual vs. collaborative) and the class/teacher.

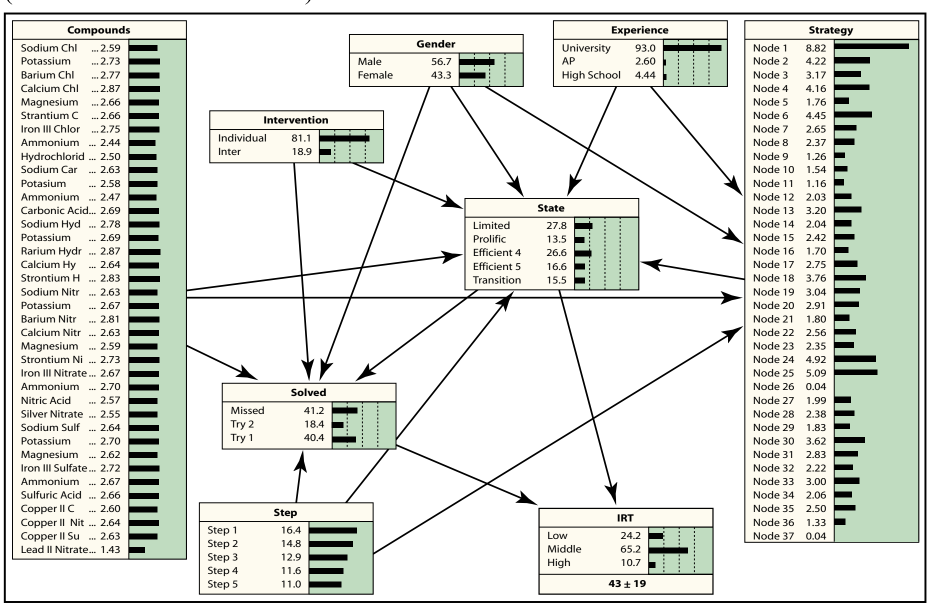

Figure 3. A Sample *Hazmat* Belief Network. Each of the variables being investigated has been divided/discretized into categories, and the bar charts indicate the proportion of the sample in each category. This figure also provides a representation of the dataset composition. For instance, there are similar numbers of males and females, but there are more university students than high school students.

Starting from the left, **Compound** represents the unknown in the case that is being solved. **The Compound → Solved** dependency acknowledges the relative difficulty of the cases given the nature of the compounds (acids, bases, salts, flame test +/-). There is also a **Compound → Strategy** association (ANN node, 1-36 from Figure 1) as would be expected as flame test negative compounds require more extensive testing than flame test positive compounds. The **Compound → State** (HMM State) link reflects the correlation between certain HMM hidden states and different compounds. In this figure, the *limited* value equals State 1 in Figure 2, *prolific* = State 3, *transition* = State 2 and *efficient_4* and *efficient_5* = States 4 and 5 respectively. This association between Strategy and State is identified from the HMM emission matrix and confirmed by X^2 analysis.

The **Intervention** variable indicates whether the problem was solved by an individual or through an intervention, which here is placing the students in collaborative groups. We have previously shown that students working in groups stabilized their strategies more rapidly than did individuals, solved a greater proportion of the problems, and used different approaches [17]. These dependencies are reflected in the **Intervention → Solved** and **Intervention → State** links.

Previous studies have also shown that while the overall problem solution frequency (Solved) is similar across gender there are significant gender differences in the Strategies and States used during the problem solving process that account for the **Gender → Solved** and **Gender → State** links [12].

More educationally advanced students represented by the **Experience** node, solve problems more effectively (**Experience → Solved)** and efficiently (**Experience → State)** [16]. As shown in this figure, this dataset primarily contains university students. Nevertheless, given the size of the dataset, a limited comparison can be made between university and high school students (Figure 2).

It is also possible to include a **Classroom** identifier that allows a finer granularity of classroom practices to be included.

The Step variable in the lower left corner acknowledges the changes in Strategies (**Step → Strategies)**, States (**Step → States)**, and Solved (**Step → Solved)** as students perform a series of *Hazmat* problems. As the problems are randomly delivered to students there are no links to **Compounds**.

The final variable included is **IRT** which is the estimate of overall student ability modeled by Item Response Theory analysis after students have solved a series of problems of varying difficulty. As the input data for IRT analysis is whether or not a problem was solved, IRT is closely correlated with the **Solved** variable (**IRT → Solved)**. Students with different IRT abilities stabilize their strategies at different rates and with different proportions of the States [15].

As most of the students performed between 5 and 10 *Hazmat* problems, the dataset contains performance data on subsequent problems allowing the incorporation of nodes for the predicted performance state (**Prediction**), as well as predictions as to whether or not a subsequent problem will be solved (**PSolved**). Such analyses can become quite refined given the ability to isolate particular values of the different variables. For example, in the Figure 4 it can be seen that if high ability students using an *efficient* strategy are given a difficult case, Carbonic Acid as their second case, they are likely to miss the case, but are likely to solve (**PSolved**) the next case with a good strategy (**Prediction**).

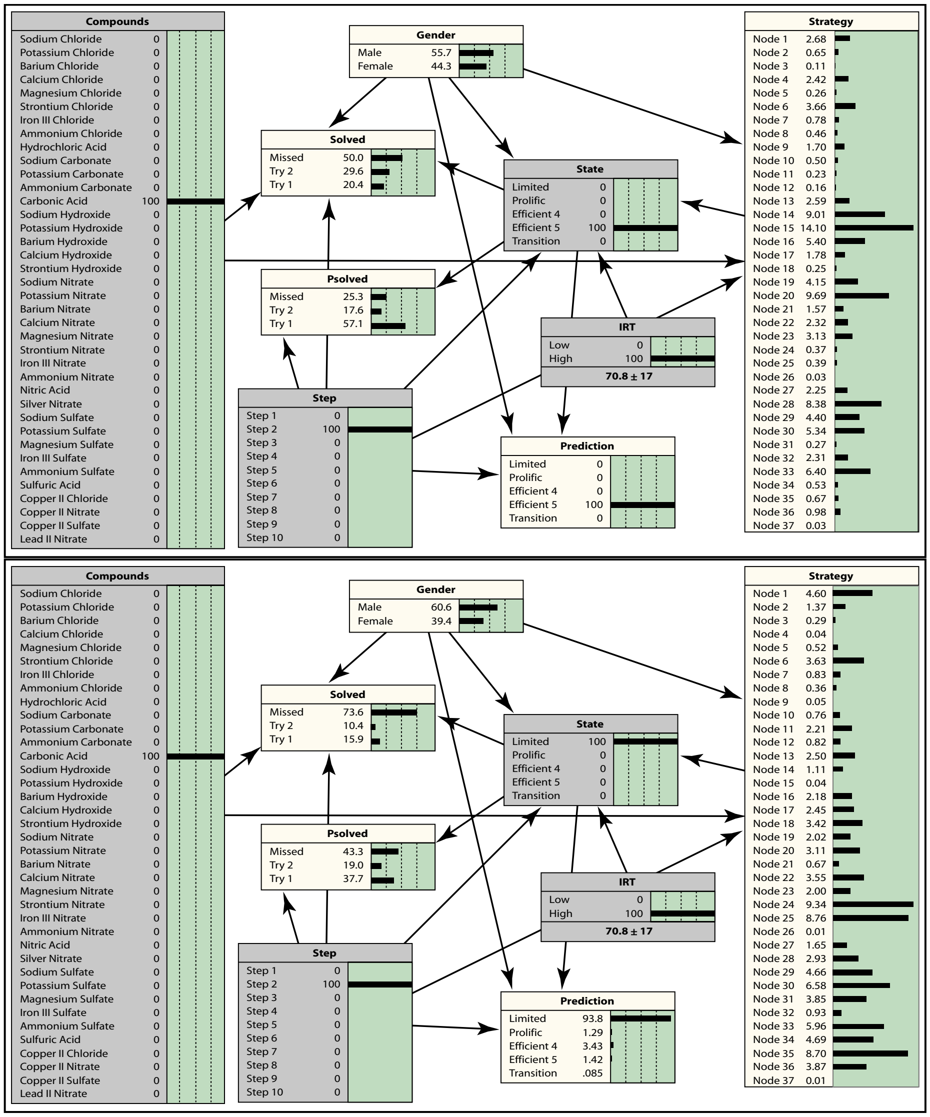

Figure 4 A Hazmat Belief Network Outcome Analysis. In this analysis the variables Compound, Step, and IRT were fixed and the State was alternated between *efficient_5* (Fig. 5a-left) and *limited* (Fig. 5b-right) to determine the outcomes for the Solved, PSolved and Prediction variables.

However, were the student to adopt a *limited* **State** under the same circumstances, then the solve rate would be lower, and the predicted solve rate on the subsequent case would also be lower (Figure 4b). This example illustrates how the output of the descriptive network could serve as a controller engine, which together with a module for sequencing tactics, could deliver pedagogical feedback between cases to improve subsequent performances. In the first example, it is unlikely that feedback would be needed, whereas in the second it would be better justified.

As part of the validation process, we analyzed the sensitivity of the findings for **State** and **Solved** with other evidence nodes (Table 1.). The goal here is to determine which of the other tests provides the best information about the State and Solved node values. A single number that is often used to best describe the sensitivity of one node to another is termed entropy which reflects the uncertainty in a probability mass. The reduction in entropy at the query node by the findings at the test nodes provides a measure of the strength of interactions.

Table 1. Sensitivity of *State* and *Solved* Nodes to Findings at Other Nodes

Findings at:	Entropy Reduction (%) at State	Findings at:	Entropy Reduction (%) at Solved
State	100	Solved	100
Strategy	17.8	Compounds	9.16
Class	7.5	Class	1.33
Step	3.2	Strategy	1.04
Solved	0.25	State	0.56
Compounds	0.23	Step	0.23
Intervention	0.2	Gender	0.04
Experience	0.07	Intervention	0.02
Gender	0.04	Experience	0.01

Such an analysis for **State** indicates that the **Strategy** contributed most to entropy reduction which makes sense as the ANN nodes constituting **Strategy** are the input symbols for the HMM analysis. Similarly, **Step** was the third highest contributor to entropy reduction, and again, this was not surprising as the HMM modeling resulting in the **State** outputs is conducted over a series of cases, i.e. a progress metric. What was less expected was that **Class** was the second highest contributor suggesting that the environmental context under which the problem solving occurred may be an important contributor to the strategy eventually used. This is consistent with earlier correlation data reported for a molecular genetics problem set [14].

A similar analysis (Table 1) of whether or not the case was **Solved** showed that **Compounds** contributed most to entropy reduction, which makes sense given the spectrum of compounds of varying difficulties [15]. The **Class** variable was the second highest contributor again pointing to the importance of the instructional environment.

To evaluate where the model is/is not functioning properly testing was performed with ~1000 student performances that were randomly removed from the dataset before the BN learning. For **State** (Table 2) there was 17% error rate with the *limited* State being the most predictable with an error rate of 9% and the *transition* state being the least predictable with an error rate of 30%. For the **Solved** variable (Table 3) the overall error was 32% and was similar for both the *Missed* and the *solved (Try_1)* values.

Table 2. Classification Error Rates for *States* When Tested with Randomly Selected Unknown Performances

Predicted					
Limited	*Prolific*	*Efficient_4*	*Efficient_5*	*Transition*	Actual
218	4	8	3	3	*Limited*
3	82	20	5	26	*Prolific*
6	6	218	5	8	*Efficient_4*
8	3	3	136	2	*Efficient_5*
4	10	7	18	90	*Transition*
Error rate = 16.96%					

Table 3. Classification Error Rates for *Solved* When Tested with Randomly Selected Unknown Performances

Predicted		
Missed	*Try_1 (Solved)*	Actual
280	152	*Missed*
135	338	*Try_1 (Solved)*
Error rate = 31.71%		

4 Discussion

These studies were motivated by the large number of statistically significant correlations we have observed between different performance metrics and a spectrum of nominal contextual variables including gender, whether the IMMEX cases were performed individually or in groups, student's academic experience, the classroom environment, and overall problem solving ability.

The BN models being developed appear consistent with prior χ^2 analyses in that the **Solved** variable is most influenced by the **Compound** being identified while the **State** variable was most influenced by the **Strategy** variable followed by **Class** and **Step**. Each of these variables would be expected to influence how the problem is framed and approached in different ways, **Compound** because of the diversity of compounds in the dataset, **Class** in that the way the problem solving is modeled for the students is likely to affect the student's own approach, and **Step** because the problem solving approaches are expected to change as experience is gained.

The most unusual finding was the only distant relationship between the **Solved** and **State** variables suggesting that these two outcomes may represent separable aspects of the problem solving process, e.g. having a correct model of a concept, and correctly applying the model, which may not only have different cognitive foundations, but may also have implications for supporting student learning.

The strategic approaches for instance, may be best represented by theories of skill acquisition. Across many observable human activities it is apparent that most individuals do not continually improve their performance. As experience is gained and students' approaches to performing the task become more routine, the incremental gains in their skills become smaller and eventually appear to stabilize. Our prior studies and those reported in Figure 1 show that on scientific problem solving tasks this skill stabilization may be accompanied by the stabilization of strategies [22] [15]. However, the strategies with which students stabilize are often not effective, indicating that, experience alone is not sufficient for some students to progress, a finding reported by others [1]. In the current study sensitivity analysis has also shown a limited dependency between whether or not the problem was solved and the approach taken during the process.

While the theory of skill acquisition helps account for the stabilization of strategic approaches, the factors influencing whether or not the problem was solved are less clear, but may relate to variables outside those currently being collected. For instance they may relate to attribution theory and the ways that students assign causality to their actions. Whether students relate their performance to internal factors, such as their own level of intelligence or to external factors such as the teacher could have significant effects on motivation, behavior and eventual outcomes. We are currently conducting parallel survey information to probe these contributions.

From the sensitivity analysis, the contribution of the classroom environment (**Class**) was one of the highest contributors to both solved and strategy being the second largest contributor for each. IMMEX is a complex tool and such complex problem solving is likely to be most effective when it is facilitated by strong instructional practices.

We are currently developing classroom practice codes that will facilitate their incorporation into our BN architectures. These codes capture instructional events or strategies that may predict students' performance and progress on IMMEX. For instance, one set of variables examines phases of IMMEX lessons on the hypothesis that particular IMMEX events -- such as the presence of an extended IMMEX "sharing" phase (during which the class discusses and reflects on a just-solved problem) -- will be positively correlated with higher performance and more effective strategy use. A second set of variables examines functions of teacher tasks (or directives) and questions; here the hypothesis is that some tasks and questions – for instance, those that promote student metacognition -- will similarly predict greater student gains.

Finally the visual interface developed through the Bayesian modeling is a valuable visualization and training tool for helping to understand the complex contributions of multiple variables to problem solving outcomes.

Supported in part by grants from the National Science Foundation (NSF-ROLE 0231995, DUE Award 0126050, ESE 9453918) and the U.S. Department of Education (R305H050052).

References

1. Ericsson, K. A. (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. Academic Medicine. 79 (10), S70-S81.
2. Fennema, E. Carpenter, T., Jacobs, V., Franke, M., and Levi, L. (1998). Gender differences in mathematical thinking. Educational Researcher, 27, 6-11.
3. Kohonen, T., (2001). Self Organizing Maps. 3rd extended edit. Springer, Berlin, Heidelberg, New York. ISBN 3-540-67921-9, ISSN 0720-678X
4. Lajoie, S.P. (2003). Transitions and Trajectories for Studies of Expertise. Educational Researcher. 32: 21-25.
5. Linacre, J.M. (2004). WINSTEPS Rasch measurement computer program. Chicago. Winsteps.com
6. Marshall, S. (1995). Schemas in Problem Solving. Cambridge University Press, ISBN: 0521430720.

7. Mayer, R. E. (2004). Should There Be A Three-Strikes Rule Against Pure Discovery Learning? American Psychologist, 59, 14-19
8. Mayer, R.E., (1998). Cognitive, Metacognitive and Motivational Aspects of Problem Solving. Instructional Science 26: 49-63.
9. Murphy, K. (2004). Hidden Markov Model (HMM) Toolbox for Matlab. Available online at: http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.
10. Olson, A., & Loucks-Horsley, S. (Eds). (2000). Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC: National Academy Press
11. Rabiner, L., (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings: IEEE, 77: 257-286.
12. Soller, A., & Lesgold, A. (2003). A Computational Approach to Analyzing Online Knowledge Sharing Interaction. Proceedings of Artificial Intelligence in Education, 2003, 253-260, Sydney Australia.
13. Soller, A., Wiebe, J., & Lesgold, A. (2002). A Machine Learning Approach to Assessing Knowledge Sharing During Collaborative Learning Activities. Proceedings of Computer Support for Collaborative Learning. Boulder, CO, pp. 128-137
14. Stevens, R. H., Ikeda, J., Casillas, A., Palacio-Cayetano, J., & Clyman, S. (1999). Artificial Neural Network-based Performance Assessments. Computers in Human Behavior, 15, 295-314.
15. 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.
16. Stevens, R., Johnson, D. F., & Soller, A. (2005). Probabilities and Predictions: Modeling the Development of Scientific Competence. Cell Biology Education, 4: 42-57. The American Society for Cell Biology.
17. Stevens, R., 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. Lester, Vicari, & Paraguaca (Eds). Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings (pp. 580-591).
18. Stevens, R.H., & Najafi, K. (1993). Artificial Neural Networks as Adjuncts for Assessing Medical Students' Problem-solving Performances on Computer-based Simulations. Computers and Biomedical Research, 26, 172-187.
19. Stevens, R.H., Ikeda, J., Casillas, A., Palacio-Cayetano, J., and S. Clyman (1999). Artificial neural network-based performance assessments. Computers in Human Behavior, 15: 295-314
20. Stevens, R.H., and Soller, (2005). Implementing a Layered Analytic Approach for Real-Time Modeling of Student's Scientific Understanding. Proceedings of the Artificial Intelligence in Education Conference, 2005.
21. VanLehn, K., Ohlsson, S., & Nason, R. (1994). Applications of Simulated Students: An Exploration. Journal of Artificial Intelligence in Education. 5(2), 135-175.
22. Vendlinski, T., Stevens, R. (2002). A Markov Model Analysis of Problem-Solving Progress and Transfer. Journal of Technology, Learning, and Assessment, 1(3).