Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor

Ron Stevens[1], Amy Soller[2], Melanie Cooper[3], Marcia Sprang[4]

[1] IMMEX Project, UCLA, 5601 W. Slauson Ave, #255, Culver City, CA. 90230
immex_ron@hotmail.com
[2] ITC-IRST, Via Sommarive 18, 38050 Povo, Trento, Italy,
soller@itc.it
[3] Department of Chemistry, Clemson University, Clemson, SC. 29634
cmelanie@clemson.edu
[4] Placentia-Yorba Linda Unified School District, 1830 N. Kellogg Drive, Anaheim, CA 92807
msprang@pylusd.k12.ca.us

Abstract. This research describes a probabilistic approach for developing predictive models of how students learn problem-solving skills in general qualitative chemistry. The goal is to use these models to apply active, real-time interventions when the learning appears less than optimal. We first use self-organizing artificial neural networks to identify the most common student strategies on the online tasks, and then apply Hidden Markov Modeling to sequences of these strategies to model learning trajectories. We have found that: strategic learning trajectories, which are consistent with theories of competence development, can be modeled with a stochastic state transition paradigm; trajectories differ across gender, collaborative groups and student ability; and, these models can be used to accurately (>80%) predict future performances. While modeling learning in chemistry developed the approach, it is applicable to many science domains where learning in a complex domain can be followed over time.

1. Introduction

Real-time modeling of how students approach and solve scientific problems is important for understanding how competence in scientific reasoning develops, and for using this understanding to improve all students' learning. Student strategies, whether successful or not, are aggregates of multiple cognitive processes [1], [2] (Anderson, 1980, Chi et al, 1988) including comprehension of the material, search for other relevant information, evaluation of the quality of the information, the drawing of appropriate inferences from the information, and the use of self-regulation processes that help keep the student on track [3], [4], [5], [6], [7]. While it is unreasonable to expect students to become domain experts, models of domain learning students should at least be expected to make significant progress marked by changes in knowledge, and strategic processing [8].

Documenting student strategies at various levels of detail can provide evidence of a student's changing understanding of the task, as well as the relative contributions of different cognitive processes to the strategy [9]. Given sufficient detail, such descriptions can provide a framework for providing feedback to the student to improve learning, particularly if the frameworks developed had predictive properties. Our long-term goal has been to develop online problem-solving systems, collectively called IMMEX (Interactive Multi-Media Exercises) to better understand how strategies are developed during scientific problem solving [10], [11]. IMMEX problem solving follows the hypothetical-deductive learning model of scientific inquiry [12], [13] where students need to frame a problem from a descriptive scenario, judge what information is relevant, plan a search strategy, gather information, and eventually reach a decision that demonstrates understanding (http://www.immex.ucla.edu). Over 100 IMMEX problem sets have been created by teams of educators, teachers, and university faculty that reflect disciplinary learning goals, and meet state and national curriculum objectives and learning standards.

In this study, the problem set we used to model strategic development is termed *Hazmat*, and provides evidence of student's ability to conduct qualitative chemical analyses (Figure 1). The problem begins with a multimedia presentation, explaining that an earthquake caused a chemical spill in the stockroom and the student's challenge is to identify the chemical. The problem space contains 22 menu items for accessing a Library of terms, the Stockroom Inventory, or for performing Physical or Chemical Testing. When the student selects a menu item, she verifies the test requested and is then shown a multimedia presentation of the test results (e.g. a precipitate forms in the liquid or the light bulb switches on suggesting an electrolytic compound). When students feel they have gathered adequate information to identify the unknown they can attempt to solve the problem. The IMMEX database collects timestamps of each student selection.

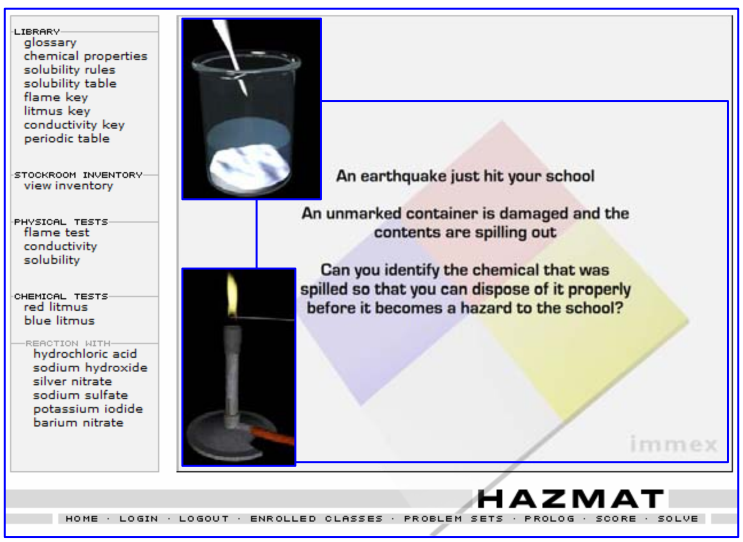

Figure 1. ***HAZMAT*** This composite screen shot of *Hazmat* illustrates the challenge to the student and shows the menu items on the left side of the screen. Also shown are two of the test items available. The item in the upper left corner shows the result of a precipitation reaction and the frame at the lower left is the result of flame testing the unknown.

To ensure that students gain adequate experience, this problem set contains 34 cases that can be performed in class, assigned as homework, or used for testing. These cases are of known difficulty from item response theory analysis (IRT [14]), helping teachers select "hard" or "easy" cases depending on their student's ability [15].

Developing learning trajectories from these sequences of intentional student actions is a two-stage process. First, the strategies used on individual cases of a problem set are identified and classified with artificial neural networks (ANN) [16], [15], [17], [18]. Then, as students solve additional problems, the sequences of strategies are modeled into performance states by Hidden Markov Modeling (HMM) [19].

1.1 Identifying Strategies with Artificial Neural Network Analysis

The most common student approaches (i.e. strategies) to solving *Hazmat* are identified with competitive, self-organizing artificial neural networks (SOM) using the student's selections of menu items as they solve the problem as input vectors [15], [17]. Self-organizing maps learn to recognize groups of similar performances in such a way that neurons near each other in the neuron layer respond to similar input vectors [20]. 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 train with between 2000-5000 performances derived from students with different ability levels (i.e. regular, honors and AP high school and university freshmen) and where each student performed at least 6 problems of the problem set. Selection criteria for the numbers of nodes, the different architectures, neighborhoods, and training parameters have been described previously [17]. The components of each strategy in this classification can be visualized for each of the 36 nodes by histograms showing the frequency of items selected (Figure 2).

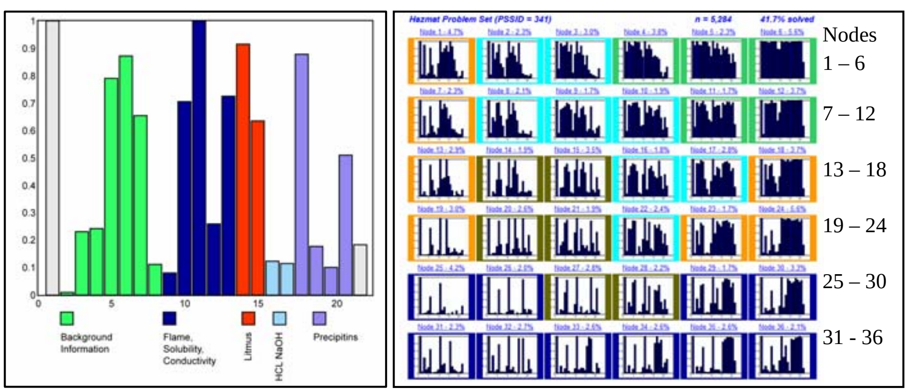

Figure 2. Sample Neural Network Nodal Analysis. A. This analysis plots the selection frequency of each item for the performances at a particular node (here, node 15). General categories of these tests are identified by the associated labels. This representation is useful for determining the characteristics of the performances at a particular node, and the relation of these performances to those of neighboring neurons. **B.** This figure shows the item selection frequencies for all 36 nodes following training with 5284 student performances.

Most strategies defined in this way consist of items that are always selected for performances at that node (i.e. those with a frequency of 1) as well as items that are ordered more variably. For instance, all Node 15 performances shown in Figure 2 A contain the items 1 (Prologue) and 11 (Flame Test). Items 5, 6, 10, 13, 14, 15 and 18 have a selection frequency of 60 - 80% and so any individual student performance would contain only some of these items. Finally, there are items with a selection frequency of 10-30%, which we regard more as background noise.

Figure 2 B is a composite ANN nodal map, which illustrates the topology generated during the self-organizing training process. Each of the 36 graphs in the matrix represents one node in the ANN, where each individual node summarizes groups of similar students problem solving performances automatically clustered together by the ANN procedure. As the neural network was trained with vectors representing the items students selected, it is not surprising that a topology developed based on the quantity of items. For instance, the upper right hand of the map (nodes 6, 12) represents strategies where a large number of tests have been ordered, whereas the lower left corner contains strategies where few tests have been ordered.

A more subtle strategic difference is where students select a large number of Reactions and Chemical Tests (items 15-21), but no longer use the Background Information (items 2-9). This strategy is represented in the lower right hand corner of Figure 2 B (nodes 29, 30, 34, 35, 36) and is characterized by extensive selection of items mainly on the right-hand side of each histogram. The lower-left hand corner and the middle of the topology map suggest more selective picking and choosing of a few, relevant items. In these cases, the SOM's show us that the students are able to solve the problem efficiently, because they know and select those items that impact their decision processes the most, and which other items are less significant.

Once ANN's are trained and the strategies represented by each node defined, then new performances can be tested on the trained neural network, and the node (strategy) that best matches the new performance can be identified. Were a student to order many tests while solving a *Hazmat* case, this performance would be classified with the nodes of the upper right hand corner of Figure 2 B, whereas a performance where few tests were ordered would be more to the left side of the ANN map. The strategies defined in this way can be aggregated by class, grade level, school, or gender, and related to other achievement and demographic measures. This classification is an observable variable that can be used for immediate feedback to the student, serve as input to a test-level scoring process, or serve as data for further research.

1.2 Hidden Markov Model Analysis of Student Progress

This section describes how we can use the ANN performance classification procedure described in the previous section to model student learning progress over multiple problem solving cases. Here students perform multiple cases in the 34-case *Hazmat* problem set, and we then classify each performance with the trained ANN (Table 1). Some sequences of performances localize to a limited portion of the ANN topology map like examples 1 or 3, suggesting only small shifts in strategy with each new performance. Other performance sequences, like example 2 show localized activity on the topology activity early in the sequence followed by large topology shifts indicating more extensive strategy shifts. Others illustrate diverse strategy shifts moving over the entire topology map (i.e. examples 4, 5).

Table 1. Student Learning Sequences. The sequence of the ANN node classifications of student performances are traced for 5 students. By mapping these sequences to the performance characteristics at each node of the trained ANN, a description of each student's progress can be generated. By examining the topology of strategy change, trajectories can be classified as *localized* (i.e. confined to a contiguous region of Figure 2 B), *progressive* (i.e. moving across the ANN topology map), or *shifting* (i.e. making larger jumps across the map)

Example	Strategy Sequence	Description	Trajectory
1	32 33 28 33 33	Limited test selections, few background resources	Localized
2	12 18 24 20 1	Many tests becoming fewer with progress.	Progressive
3	5 24 6 18	Many test selections.	Localized
4	4 22 33 33	Resource extensive going to data.	Shifting
5	4 6 14 25 30 19	Shifts between data rich and data lean strategies.	Shifting

While informative, manual inspection and mapping of nodes to strategies is a time-consuming process. One approach for dynamically, and automatically modeling this information would be to probabilistically link the strategic transitions. However, with 1296 possible transitions in a 36-neuron map, full probabilistic models would likely lack predictive power.

By using HMMs we have been able to aggregate the data and model the development and progression of generalized performance characteristics. HMM's are used to model processes that move stochastically through a series of predefined states [19]. These methods had been used successfully in previous research efforts to characterize sequences of collaborative problem solving interaction, leading us to believe that they might show promise for also understanding individual problem solving [21], [22].

In our HMMs for describing student strategy development, we postulate, from a cognitive task analysis, between 3-5 states that students may pass through as competence develops. Then, many exemplars of sequences of strategies (ANN node classifications) are repeatedly presented to the HMM modeling software to model progress. These models are defined by a transition matrix that shows the probability of transiting from one state to another and an emission matrix that relates each state back to the ANN nodes that best represent that state. (Murphy, K. http://www.ai.mit.edu/~murphyk/Software/HMM/hmm.html). Recall from the previous section that each of these nodes characterizes a particular problem solving strategy. The transitions between the 5 states describe the probability of students transitioning between problem solving strategies as they perform a series of IMMEX cases. While the emission matrices associated with each state provides a link between student performances (ANN node classification) and progress (HMM states), the transition matrix (describing the probability of moving from each state in the HMM to each other state) can be used for analyzing / predicting subsequent performances.

Both of these features are shown in Figure 3 with the transitions between the different states in the center, and the ANN nodes representing each state at the periphery. States 1, 4, and 5 appear to be absorbing states as these strategies once used are likely to be used again. In contrast, students adopting State 2 and 3 strategies are less likely to persist with those states but are more likely to transit to another state. When the emission matrix of each state was overlaid on the 6 x 6 neural network grid, each state (Figure 4), represented topology regions of the neural network that were often contiguous (with the exception of State 4).

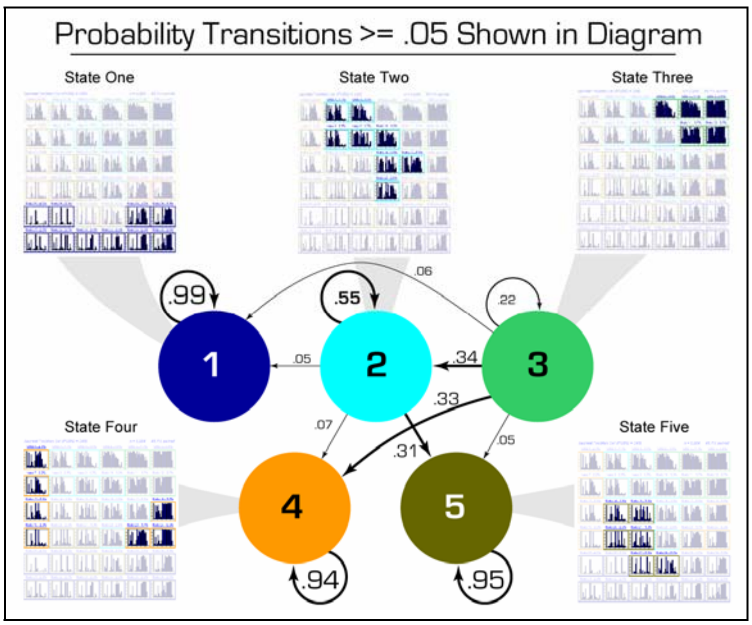

Figure 3. Mapping the HMM Emission and Transition Matrices to Artificial Neural Network Classifications. The five states comprising the HMM for *Hazmat* are indicated by the central circles with the transitions between the states shown by the arrows. Surrounding the states are the artificial neural network nodes most closely associated with each state.

2. Results

As we wish to use the HMM to determine how students strategic reasoning changes with time, we performed initial validation studies to determine 1) how the state distribution changes with the number of cases performed, 2) whether these changes reflect learning progress, and, 3) whether the changes over time 'make sense' from the perspective of novice/expert cognitive differences.

The overall solution frequency for the Hazmat dataset (N= 7630 performances) was 56%, and when mapped to the HMM states provided the following quantitative and qualitative descriptions:

- State 1 – 55% solution frequency showing variable numbers of test items and little use of Background Information;
- State 2 – 60% solution frequency showing equal usage of Background Information as well as action items; little use of precipitation reactions.
- State 3 – 45% solution frequency with nearly all items being selected.
- State 4 – 54% solution frequency with many test items and limited use of Background Information.
- State 5 – 70% solution frequency with few items selected Litmus test and Flame tests uniformly present.

We next profiled the states for the dynamics of state changes, and possible gender and group vs. individual performance differences.

Dynamics of State Changes – Across 7 *Hazmat* performances the solved rate increased from 53% (case 1) to 62% (case 5) (Pearson χ^2 =15.5, p=.008) and this was accompanied by corresponding state changes (Figure 4). These longitudinal changes were characterized by a decrease in the proportions of States 1 and 3 performances and an increase and then decrease in State 2 performances and a general increase in State 5 (with the highest solution frequency).

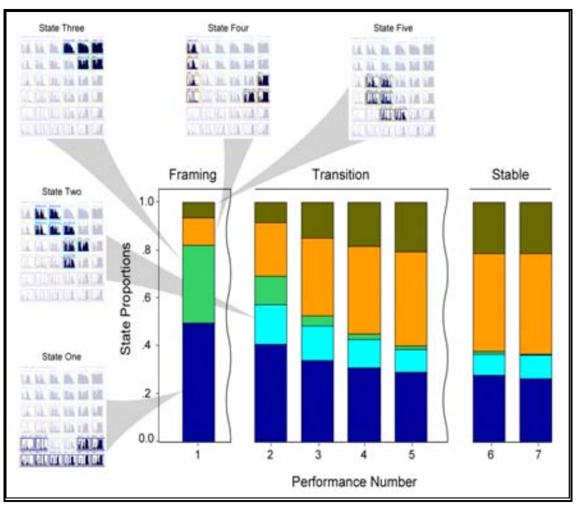

Figure 4. Dynamics of HMM State Distributions with Experience and Across Classrooms. The bar chart tracks the changes in all student strategy states (n=7196) across seven *Hazmat* performances. Mini-frames of the strategies in each state are shown for reference.

Group vs. Individual Performance - In some settings the students worked on the cases in teams of 2-3 rather than individually. Group performance significantly increased the solution frequency from a 51% solve rate for individuals to 63% for the students in groups. Strategically, the most notable differences were the maintenance of State 1 as the dominant state, the nearly complete lack of performances in States 2 and 3, and the more rapid adoption of State 4 performances by the groups (Figure 5). In addition, the groups stabilized their performances faster, changing little after the third performance whereas males and females stabilized only after performance 5. This makes sense because states 2 and 3 represent transitional phases that students pass through as they develop competence. Collaborative learners may spend less time in these phases if group interaction indeed helps students see multiple perspectives and reconcile different viewpoints to better able to overcome these impasses [23].

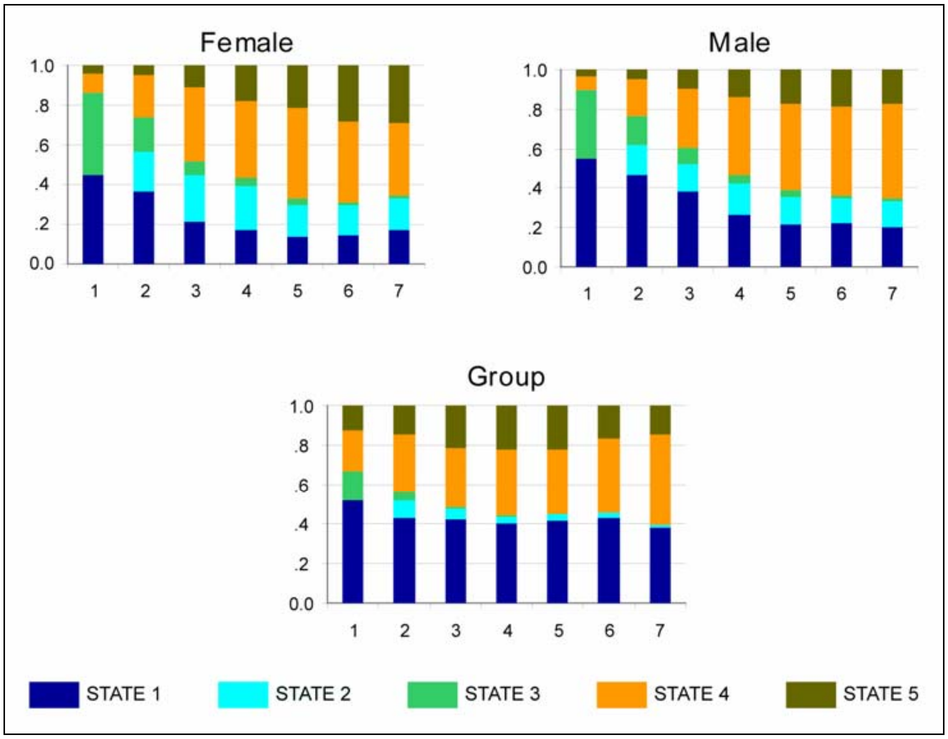

Figure 5. State Distributions for Individuals and Groups

Also, shown in Figure 5 are the differences in the state distribution of performances across males and females ((Pearson $\chi^2 = 31.2$, $p<0.000$). While there was a steady reduction in State 1 performances for both groups, the females entered State 2 more rapidly and exited more rapidly to State 5. These differences became non-significant at the stable phase of the trajectories (performances 6 and 7). Thus males and females have different learning trajectories but appear to arrive at similar strategy states.

Ability and State Transitions - Learning trajectories were then developed according to student ability as determined by IRT. For these studies, students were grouped into high (person measure = 72-99, n = 1300), medium (person measure 50-72, n = 4336) and low (person measure 20-50, n = 1994) abilities. As expected from the nature of IRT, the percentage solved rate correlated with student ability. What was less expected was that when the solved rate by ability was examined for the sequence of performances, the students with the lowest ability had not only the highest solved rate on the first performance, but also one that was significantly better than the highest ability students (57% vs 44% n = 866, $p < 0.00$). Predictably, this was rapidly reversed on subsequent cases. To better understand these framing differences a cross-tabulation analysis was conducted between student ability and neural network nodal classifications on the first performances. This analysis highlighted nodes 3, 4, 18, 19, 25, 26, and 31 as having the highest residuals for the low ability students, and nodes 5, 6, 12, and 17 for the highest ability students. From this data, it appeared that the higher ability students chose to more thoroughly explore the problem space on their first performance, to the detriment of their solution frequency, but took advantage of this knowledge on subsequent performances to improve their strategies. These improvements during the transition and stabilization stages include increased use of State 5 performances, and decreased use of States 1 and 4; i.e. they become both more efficient and effective

Predicting Future Student Strategies - An additional advantage of a HMM is that predictions can be made regarding the student's learning trajectory. The prediction accuracy was tested in the following way. First, a 'true' mapping of each node and the corresponding state was conducted for each performance of a performance sequence. For each step of each sequence, i.e. going from performance 2 to 3, or 3 to 4, or 4 to 5, the posterior state probabilities of the emission sequence (ANN nodes) were calculated to give the probability that the HMM is in a particular state when it generated a symbol in the sequence, given that the sequence was emitted. For instance, ANN nodal sequence [6 18 1] mapped to HMM states (3 4 4). Then, this 'true' value is compared with the most likely value obtained when the last sequence value was substituted by each of the 36 possible emissions representing the 36 ANN nodes describing the student strategies. For instance, the HMM calculated the likelihood of the emission sequences, [6 18 X] in each case where X = 1 to 36. The most likely emission value for X (the student's most likely next strategy) was given by the sequence with the highest probability of occurrence, given the trained HMM. The student's most likely next performance state was then given by the state with the maximum likelihood for that sequence.

Comparing the 'true' state values with the predicted values estimated the predictive accuracy of the model at nearly 90% (Table 2). As the performance sequence increased, the prediction rate also increased, most likely reflecting that by performances 4, 5 and 6, students are repeatedly using similar strategies.

Table 2. Prediction of Future Performances

Performance #	1	2	3	4	5	6
% Correct Predictions	67	75	83	88	86	91

3. Discussion

The goal of this study was to explore the use of HMMs to begin to model how students gain competence in domain-specific problem solving. The idea of 'learning trajectories' is useful when thinking about how students progress on the road to competence [24]. These trajectories are developed from the different ways that novices and experts think and perform in a domain, and can be thought of as defining stages of understanding of a domain or discipline [4]. During early learning, students' domain knowledge is limited and fragmented, the terminology is uncertain and it is difficult for them to know how to properly frame problems. In our models, this first strategic stage is best represented by State 3 where students extensively explore the problem space and select many of the available items. As expected, the solved rate for such a strategy was poor. This approach is characteristic of surface level strategies or those built from situational (and perhaps inaccurate) experiences. From the transition matrix in Figure 4, State 3 is not an absorbing state and most students move from this strategy type on subsequent performances.

With experience, the student's knowledge base becomes qualitatively more structured and quantitatively deeper and this is reflected in the way competent students, or experts approach and solve difficult domain-related problems. In our model States 2 and 4 would best represent the beginning of this stage of understanding. State 2 consists of an equal selection of background information and test information, suggesting a lack of familiarity of the nature of the data being observed. State 4 on the other hand shows little/no selection of background information but still extensive and non-discriminating test item selection. Whereas State 2 is a transition state, State 4 is an absorbing state - perhaps one warranting intervention for students who persist with strategies represented by this state.

Once competence is developed students would be expected to employ both effective and efficient strategies. These are most clearly shown by our States 1 and 5. These states show an interesting dichotomy in that they are differentially represented in the male and female populations with males having a higher than expected number of State 1 strategies and females higher than expected State 5 strategies.

The solution frequencies at each state provide an interesting view of progress. For instance, if we compare the earlier differences in solution frequencies with the most likely state transitions from the matrix shown in Figure 4, we see that most of the students who enter State 3, having the lowest problem solving rate (45%), will transit either to State 2 or 4. Those students who transit from State 3 to 2 will show on average a 15% performance increase (from 45% to 60%) and those students who transit from States 3 to 4 will show on average a 9% performance increase (from 45% to 54%). The transition matrix also shows that students who are performing in State 2 (with a 60% solve rate) will tend to either stay in that state, or transit to State 5, showing a 10% performance increase (from 60% to 70%). This analysis shows that students' performance increases as they solve science inquiry problems through the IMMEX Interactive Learning Environment, and that by using ANN and HMM methods, we are able to track and understand their progress.

When given enough data about student's previous performances, our HMM models performed at over 90% accuracy when tasked to predict the most likely problem solving strategy the student will apply next. Knowing whether or not a student is likely to continue to use an inefficient problem solving strategy allows us to determine whether or not the student is likely to need help in the near future. Perhaps more interestingly, however, is the possibility that knowing the distribution of students' problem solving strategies and their most likely future behaviors may allow us to strategically construct collaborative learning groups containing heterogeneous combinations of various behaviors such that intervention by a human instructor is required less often [25].

Finally, our studies provide some information on the effects of collaborative learning when students perform the cases. In particular, collaborative problem solving appeared to reduce the use of strategies in States 2 and 3, which are the most transitory states. In this regard, one effect of the collaboration may be to help groups more rapidly establish stable patterns of problem solving. A question of interest would be whether these states persist once students again engaged in individual problem solving.

Supported in part by grants from the National Science Foundation (ROLE 0231995, DUE 0126050, ESE 9453918), the PT[3] Program of the U.S. Department of Education (P342A-990532), and the Howard Hughes Medical Institute Precollege Initiative.

References

1. Anderson, J.D.,(1980). Cognitive psychology and its implications. San Francisco: W.H. Freeman.
2. Chi, M. T. H., Glaser, R., and Farr, M.J. (eds.), (1988). The Nature of Expertise, Hillsdale, Lawrence Erlbaum, pp 129-152.
3. Chi, M. T. H., Bassok, M., Lewis, M. W., Reinmann, P., and Glaser, R. (1989). Self-Explanations: how students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182.
4. VanLehn, K., (1996). Cognitive Skill Acquisition. Annu. Rev. Psychol 47: 513-539.
5. Schunn, C.D., and Anderson, J.R. (2002). The generality/specificity of expertise in scientific reasoning. Cognitive Science.
6. Corbett, A. T. & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4, 253-278.
7. Schunn, C.D., Lovett, M.C., and Reder, L.M. (2001). Awareness and working memory in strategy adaptivity. Memory & Cognition, 29(2); 254-266.
8. Haider, H., and Frensch, P.A. (1996). The role of information reduction in skill acquisition. Cognitive Psychology 30: 304-337.
9. Alexander, P., (2003). The development of expertise: the journey from acclimation to proficiency. Educational Researcher, 32: (8), 10-14.

10. 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.

11. Underdahl, J., Palacio-Cayetano, J., and Stevens, R., (2001). Practice makes perfect: assessing and enhancing knowledge and problem-solving skills with IMMEX software. *Learning and Leading with Technology*. 28: 26-31.

12. Lawson, A.E. (1995). *Science Teaching and the Development of Thinking.* Wadsworth Publishing Company, Belmont, California

13. 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.

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

15. Stevens, R.H., and Najafi K. (1993). Artificial neural networks as adjuncts for assessing Medical students' problem-solving performances on computer-based simulations. *Computers and Biomedical Research* 26(2), 172-187.

16. Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. Cambridge, MA: MIT Press

17. Stevens, R., Wang, P., Lopo, A. (1996). Artificial neural networks can distinguish novice and expert strategies during complex problem solving. *JAMIA* vol. 3 Number 2 p 131-138

18. Casillas, A.M., Clyman, S.G., Fan, Y.V., and Stevens, R.H. (1999). Exploring alternative models of complex patient management with artificial neural networks. Advances in Health Sciences Education 1: 1-19, 1999.

19. Rabiner, L., (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc. IEEE, 77: 257-286.

20. Kohonen, T., 2001. Self Organizing Maps. 3rd extended edit. Springer, Berlin, Heidelberg, New York.

21. Soller, A. (2004). Understanding knowledge sharing breakdowns: A meeting of the quantitative and qualitative minds. Journal of Computer Assisted Learning (in press).

22. Soller, A., and Lesgold, A. (2003). A computational approach to analyzing online knowledge sharing interaction. Proceedings of Artificial Intelligence in Education, 2003. Australia, 253-260.

23. Lesgold, A., Katz, S., Greenberg, L., Hughes, E., & Eggan, G. (1992). Extensions of intelligent tutoring paradigms to support collaborative learning. In S. Dijkstra, H. Krammer, & J. van Merrienboer (Eds.), *Instructional Models in Computer-Based Learning Environments*. Berlin: Springer-Verlag, 291-311.

24. Lajoie, S.P. (2003). Transitions and trajectories for studies of expertise. Educational Researcher, 32: 21-25.

25. Giordani, A., & Soller, A. (2004). Strategic Collaboration Support in a Web-based Scientific Inquiry Environment. European Conference on Artificial Intelligence, “Workshop on Artificial Intelligence in Computer Supported Collaborative Learning”, Valencia, Spain.