MODELING THE DEVELOPMENT OF SCIENTIFIC COMPETENCE IN CHEMISTRY

Ron Stevens¹, Amy Soller², Melanie Cooper³, Marcia Sprang⁴

¹ IMMEX Project, UCLA, 5601 W. Slauson Ave, #255, Culver City, CA. 90230
² Institute for Defense Analysis, 4850 Mark Center Drive, Alexandria VA, 22311
³ Department of Chemistry, Clemson University, Clemson, SC. 29634
⁴ Placentia-Yorba Linda Unified School District, 1830 N. Kellogg Drive, Anaheim, CA 92807

ABSTRACT.
We have developed models of how strategies are constructed and retained as male and female high school and university students gain experience in solving online qualitative chemical analysis problems. The most common strategies that students used when solving online problem-solving simulations were first classified using self-organizing artificial neural networks. This resulted in strategy maps detailing the spectrum of problem solving strategy and highlighting the qualitative and quantitative differences among these approaches. Next, learning trajectories were developed by Hidden Markov Modeling of sequences of performances, which stochastically described student’s progress in understanding chemistry. We have found that students quickly establish preferred strategic approaches and that these stabilized strategic approaches are re-used up to four months later to solve additional cases. While males and females correctly identified the same percentage of unknown compounds, the proportions of strategies developed and retained were significantly different across gender groupings.

KEYWORDS.
Neural networks, Markov modeling, strategy, problem solving.

1. INTRODUCTION
Modeling how students approach and solve scientific problems is important for understanding how 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 [2], [5] 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 [4], [26], [18], [6], [19].

Modeling student strategies at various levels of detail and at different points in time as they gain experience can provide evidence of a student’s changing understanding of a task, as well as the relative contributions of different cognitive processes to the strategy [1]. Given sufficient detail, such models could extend our understanding of how gender, and other student characteristics differentially influence performance and participation in complex problem-solving environments [8]. If the models also had predictive properties, they could also provide a framework for directing feedback to improve learning.

To approach these questions we have developed online problem-solving systems, collectively called IMMEX (Interactive Multi-Media Exercises), to better understand how rapidly strategies develop, how long these strategies persist, and what influences strategy adoption and persistence. [24], [25]. IMMEX problem solving follows the hypothetical-deductive learning model of scientific inquiry [12], [15] 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. 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.

![Figure 1. HAZMAT](image)

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.

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) [23], [17], [22], [3]. Then, as students solve additional problems, the sequences of strategies are modeled into performance states by Hidden Markov Modeling (HMM) [16].

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 [23], [22]. 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 [10]. 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. The strategy components in this classification can be visualized for each of the 36 nodes by histograms showing the frequency of items selected (Figure 2). Strategies so defined consist of items that are always selected for performances at that node (i.e. with a frequency of 1) as well as items...
ordered more variably. For instance, all Node 15 performances shown in Figure 2 A contain 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 as background noise.

![Figure 2](image)

**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.

Figure 2 B is a composite ANN nodal map of the topology of performances generated during the self-organizing training process. Each of the 36 graphs in the matrix represents one node in the ANN, where similar students problem solving performances become 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 shown in the lower right hand corner of Figure 2 B (nodes 29, 30, 34, 35, 36) where students select a large number of Reactions and Chemical Tests (items 15-21), but no longer use the Background Information (items 2-9). 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 and reported. 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 on the left of the ANN map. The strategies can be aggregated by class, grade level, school, or gender, and related to other achievement and demographic measures.

### 1.2 Hidden Markov Model Analysis of Student Progress

On their own, artificial neural network analyses provide point-in-time snapshots of student's problem solving. While useful for providing within task feedback to students and teachers, the feedback potential could be amplified if across-task models could be developed, particularly if they had predictive properties. This section describes how we can use the ANN performance classification procedure described above to model student learning progress over multiple problem solving episodes. 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 performance sequence followed by large topology shifts indicating more extensive strategy revisions. 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 4 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
<table>
<thead>
<tr>
<th>Example</th>
<th>Node Sequence</th>
<th>Description</th>
<th>Trajectory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32 33 28 33 33</td>
<td>Limited test selections, few background resources</td>
<td>Localized</td>
</tr>
<tr>
<td>2</td>
<td>12 18 24 20 1</td>
<td>Many tests becoming fewer with progress.</td>
<td>Progressive</td>
</tr>
<tr>
<td>3</td>
<td>5 24 6 18</td>
<td>Many test selections.</td>
<td>Localized</td>
</tr>
<tr>
<td>4</td>
<td>4 22 33 33</td>
<td>Background shifting to data.</td>
<td>Shifting</td>
</tr>
<tr>
<td>5</td>
<td>4 6 14 25 30 19</td>
<td>Shifts between data rich and data lean strategies.</td>
<td>Shifting</td>
</tr>
</tbody>
</table>

While informative, manual mapping of nodes to sequences of strategies is time-consuming. One approach to dynamically, and automatically model 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 HMM's 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, which are not directly observed, but are associated with a probability distribution function [3]. 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 [20], [21].

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 they develop problem-solving experience. 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 transitioning 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). Such a mapping between nodes and states is shown in Figure 3, where the nodes associated with each state are overlaid on the 6 x 6 neural network grid.

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. This is shown in Table 2.

Table 2. HMM Transition Matrix. This matrix shows the likelihood of transitioning from one state to another. By looking along the diagonal (bold), States 1, 4, and 5 appear stable suggesting that once adopted, students are likely to continue to use them. In contrast, students adopting State 2 and 3 strategies are less likely to persist with those approaches but are more likely to adopt other strategies (gray boxes).

<table>
<thead>
<tr>
<th>From State</th>
<th>To State:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0.994</td>
</tr>
<tr>
<td>2</td>
<td>0.053</td>
</tr>
<tr>
<td>3</td>
<td>0.056</td>
</tr>
<tr>
<td>4</td>
<td>0.029</td>
</tr>
<tr>
<td>5</td>
<td>0.022</td>
</tr>
</tbody>
</table>

2. RESULTS

HMM State Properties - The solution frequency for the Hazmat cases (n=5258) was 51.9% with males and females performing similarly (males=53% solved, females = 51% solved, Pearson $\chi^2 = 2.8$, p=.095). The solution frequency was different for the different HMM states, which had the following properties:

- State 1 had a 55% solution frequency and showed little use of Background Information and
variable use of test items.

- State 2 had a 60% solution frequency and showed equal usage of Background Information and test items; there was little use of precipitation reactions.
- State 3 had a 45% solution frequency and the performances exhibited extensive testing.
- State 4 had a 54% solution frequency with limited use of Background Information but many test items.
- State 5 had a 70% solution frequency and relatively few test items were selected. The Litmus and Flame tests were uniformly present.

**Dynamics of State Changes With Experience** – We have previously shown that when performing a series of cases, students shift their strategic approaches over the first 3-4 performances and then stabilize as they develop strategies they are comfortable with [13]. An example of these dynamics is shown in Figure 3.

Across 7 Hazmat performances the solved rate increased from 53% (case 1) to 62% (case 5) and this was accompanied by corresponding state changes 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).

**Gender Influences** – The literature has suggested that in scientific problem solving, females as a group, are more cautious, conservative, and consistent in following the rules, while males take greater risks and respect the rules less [7]. Within the context of IMMEX problem solving, this could be differentially reflected in the quantity of tests chosen. For example, selecting fewer tests may reflect a more risky strategy, while selecting more tests could suggest a more conservative approach. Consistent with this hypothesis, on the first case of the problem set, there was a significant difference in state usage with males preferring State 1 approaches and females preferring State 3 approaches (Pearson $\chi^2 = 13.54, P<0.004$). State 1 consists of Nodes 25, 26, and 29-36 which represent performances often of the limited data type. State 3 is just the opposite where many of the available background and test items are selected. During subsequent cases, there was a steady reduction in States 1 and 3, a transient appearance of State 2 performances and then the emergence of States 4 and 5 performances. Males and females used State 4 strategies equally, while more females progressed to State 5 strategies. (Pearson $\chi^2 = 31.2, p<0.000$).

**Strategy Stabilization and Persistence** - With a smaller set of advanced placement chemistry students (3 classrooms from the same teacher, 79 students) we then explored the short and long-term stability of student’s strategies and the influences of gender. In a standard classroom environment students first performed 5 Hazmat problems to refine and stabilize their strategies. Then, one week (short-term) and 15 weeks later (long-term) students were asked to solve additional Hazmat cases.

At the end of the required first-set of performances (# 1-5), the proportions of the five strategy states and the solution frequencies had stabilized. As expected, State 3 approaches were preferred on the early problem solving performances, and these decreased over time with the emergence of States 2, 4, and 5. The proportion of State 1 strategies in this subset of students was lower than the overall population, and this was most likely due to the more controlled classroom nature of this assignment that reduced guessing.

![Figure 3. Persistence of stabilized states. The bar chart tracks the changes in all student strategy states across seven Hazmat performances. Mini-frames of the strategies in each state are shown for reference.](image)
One week, and fifteen weeks later the students were asked to perform an additional 3 Hazmat cases in class. The state distributions of the performances at both time intervals were not significantly different from those established after the initial training. It is also interesting that the solution frequency also did not change. Combined, these data indicate that students adopt a preferential approach to solving Hazmat after relatively few cases (4-5) and, as a group, they continue to use these strategies when presented with repeat cases, even after prolonged periods of time.

The performances were then separated by gender and the state distributions were re-plotted. As shown in Figure 5, both male and female students appeared to have stabilized their strategic approaches by the fifth performance, but the state distributions were significantly different, with females preferring the approaches represented by State 5 whereas the males preferred State 4 approaches.

3. DISCUSSION

The goal of this study was to track the development and persistence of problem solving strategies as students become competent in qualitative chemical problem solving, and to determine what role, if any, gender played in strategic learning and progress.

The data in Figures 4 and 5 along with the HMM transition matrix, suggest the following model for how students gain proficiency in solving qualitative analysis problems. Most students approached the first Hazmat case by selecting either an extensive (State 3), or limited (State 1) amount of information. State 1,
with an average solution rate, is an absorbing state meaning that once this limited approach is adopted by students, they are unlikely to change from it. State 1 strategies are also used significantly more frequently by males than females. The transitions among States 2-5 are more complex than State 1 and are shown in Figure 6. State 3 is an early set of approaches during Hazmat problem solving, and is also a transition state. Students who first employ these approaches are likely to switch to States 2 or 4, and increase their likelihood of solving the cases from 45% to 60% and 54% respectively.

The main difference between States 2 and 4 is that there is both test and background information being accessed in the State 2 approaches whereas State 4 are primarily data driven approaches. State 4 is interesting in several regards. First, it differs from the other states in that the strategies represented are located at distant points on the ANN topology map, whereas the nodes comprising the other states are contiguous. As State 4 strategies appear with problem-solving experience and after the students have had an opportunity to explore the problem space, they should begin to reflect their understanding of qualitative chemical analysis. The State 4 strategies represented by ANN nodes 1, 7, 13 and 19 on the left hand of the topology map are very appropriate for the set of cases in Hazmat that involve flame test positive compounds. The strategies defined by ANN nodes 18, 23 and 24 are appropriate for flame test negative compounds where more extensive testing for both the anion and cation is required. This would suggest that students using State 4 strategic approaches have mentally partitioned the Hazmat problem space into two groups of strategies, depending on the outcome of the initial flame test.

![Figure 6. A model for the evolution of Hazmat problem solving. This model, based on the HMM transition matrix and the data in Figures 4 and 5, shows the primary learning trajectories for Hazmat. The open arrows show the trajectories used more frequently by females, and the solid arrows show those used more by males.](image)

State 5 are also complex strategies which from the transition matrix in Table 2, emerge from State 2 strategies by a further reduction in the use of background resources. State 5 approaches appear late in problem solving sequences, have the highest solution frequencies and are approaches that work well with both flame test positive and negative compounds. They are also strategies used more frequently by females than by males.

While there is differential use of State 1 and State 5 strategies by males and females, it is interesting that there were no gender differences in the solved rate for any of the states. This makes it unlikely that one approach is more efficient/effective on a gender basis, and suggests that the differential approaches are more of a preference issue, with females preferring a more gradual evolution of their strategic approaches, and males more limited and/or risky strategies. This is consistent with results in mathematics problem solvingNevertheless, once adopted, each of the approaches appears equally persistent.

Supported in part by grants from the National Science Foundation (ROLE 0231995, DUE 0126050, HRD-0429156), the PT3 Program of the U.S. Department of Education (P342A-990532), and the Howard Hughes Medical Institute Precollege Initiative.
REFERENCES


