J·T·L·A

The Journal of Technology, Learning, and Assessment

Volume 1, Number 3 · June 2002

Assessing Student Problem-Solving Skills With Complex Computer-Based Tasks

Terry Vendlinski & Ron Stevens

www.jtla.org

A publication of the Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College

J·T·L·A

Volume 1, Number 3

Assessing Student Problem-Solving Skills With Complex Computer-Based Tasks
(formerly *A Markov Model Analysis of Problem-Solving Progress and Transfer*)

Terry Vendlinski and Ron Stevens

Editor: Michael Russell
russelmh@bc.edu
Technology and Assessment Study Collaborative
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Copy Editor: Kathleen O'Connor
Design and Layout: Thomas Hoffmann

JTLA is a free on-line journal, published by the Technology and Assessment Study Collaborative, Caroline A. & Peter S. Lynch School of Education, Boston College.

Preferred citation:

Vendlinski, T. & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment, 1*(3). Available from http://www.jtla.org.

Abstract:

Valid formative assessment is an essential element in improving both student learning and the professional development of educators. Various shortcomings in common assessment modalities, however, hinder our ability to make and evaluate such formative decisions. The diffusion of computer technology into American classrooms offers new opportunities to evaluate student learning and a rich, new source of data upon which to make inferences about the formative interventions that will improve learning. The path from data to inference, however, requires appropriate methodologies that can fully exploit the data without discarding or oversimplifying the behavioral complexity of student activity. This study used IMMEX™, a computerized simulation and problem-solving tool, along with artificial neural networks as pattern recognizers to identify the common types of strategies high school chemistry students used to solve qualitative chemistry problems. Then, based on the calculated probabilities that students would transition between these strategy types over time, Markov hidden chain analysis allowed us to develop a model of the capacity of the current curriculum to produce students able to apply chemistry content to a real-world problem.

Assessing Student Problem-Solving Skills With Complex Computer-Based Tasks

Terry Vendlinski and Ron Stevens
UCLA / IMMEX Lab

Background

Valid formative assessment is essential to the advancement of student learning and the development of pedagogical content knowledge in teachers (Nathan, Koedinger, & Martha, 2001). Most currently accepted pencil-and-paper standardized tests, however, are not designed as formative assessment tools (Bransford, Brown, & Cocking, 1999; American Educational Research Association, 2000), and many performance-based assessments suffer from validity (Barton, 1999), pedagogical (Lowyck & Poysa, 2001), logistic, time and cost problems (Quellmalz, Schank, Hinojosa, & Padilla, 1999). Moreover, recent thinking in the field of educational assessment suggests that formative assessments must focus less on how closely student responses match a pre-determined model and more on the competency of the performance as a whole (Pellegrino, Chudowsky, & Glaser, 2001). So, while the unstructured nature of such student responses makes the evaluation of these types of performances difficult, the need for such evaluations is likely to increase. As computer hardware becomes cheaper, connectivity easier, and software development more rapid, computerized learning and assessment simulations arguably will become ubiquitous throughout the American educational system. With the appropriate methodologies to analyze and fully exploit the rich source of data from performances on these types of simulations, new ways of informing formative pedagogical interventions in a timely and valid manner seem possible.

The Interactive Multi-media Exercises (IMMEX™) software is a web-based problem authoring, presentation, and assessment tool that allows teachers to develop and present domain specific simulations to their students. The software has proven effective because it allows simulation authors to combine a real-world problem with specific tests and reference items that may yield information about that problem, and it presents students with the opportunity to develop and test hypotheses about a solution to the given challenge in a realistic setting. Authors may also add reference items that students can access in their attempt to solve the problem. Moreover, the authoring tools in IMMEX allow educators to tailor the content of existing simulations to the specific curricular objectives and particular classroom and student contextual variables an instructor feels are important. For example, teachers may delete certain reference items for Advanced Placement (AP) students because students are required to memorize these items for the AP test. By allowing authors to change the results of specific tests, these same authoring tools also allow teachers to easily develop different versions (cases) of each problem. For example, the problem set in this study contains cases like sodium hydroxide, potassium carbonate, and lithium nitrate. The cases in a medical school problem set might be different diseases requiring diagnosis. Multiple cases allow students numerous opportunities to apply their knowledge in similar but not identical situations or to attempt problems with different degrees of difficulty. Different cases also allow teachers to monitor changes in student performance over time. Although the *number* of informational items available to the student does not change from case to case, the content knowledge required to interpret various test results can vary widely between cases. Consequently, a single problem set containing multiple cases can be used to assess students of diverse ability levels. While standard Item Response analysis has produced good models of case difficulty both in the problem set discussed in this paper as well as in other IMMEX problem sets, most teachers currently choose particular problem sets and the specific cases their students will attempt to solve more subjectively. For this study, the teacher developed 23 cases of an IMMEX qualitative chemistry problem set as a tool to assess how well her students could apply the concepts taught in first-year high school chemistry.

After the initial presentation of the problem statement (see Figure 1), students solve IMMEX cases by accessing as much of the available information as they feel necessary and then selecting an answer from a list of possible answers or by typing in their solution. While students can develop and test all their hypotheses by accessing information in any sequence they choose, the verification or rejection of each hypothesis usually relies on how well the student interprets individual test results. An example of the information a student would be expected to interpret is also shown in Figure 1. In this case, a student should be able to retain or reject a hypothesized unknown based on these results.

Figure 1

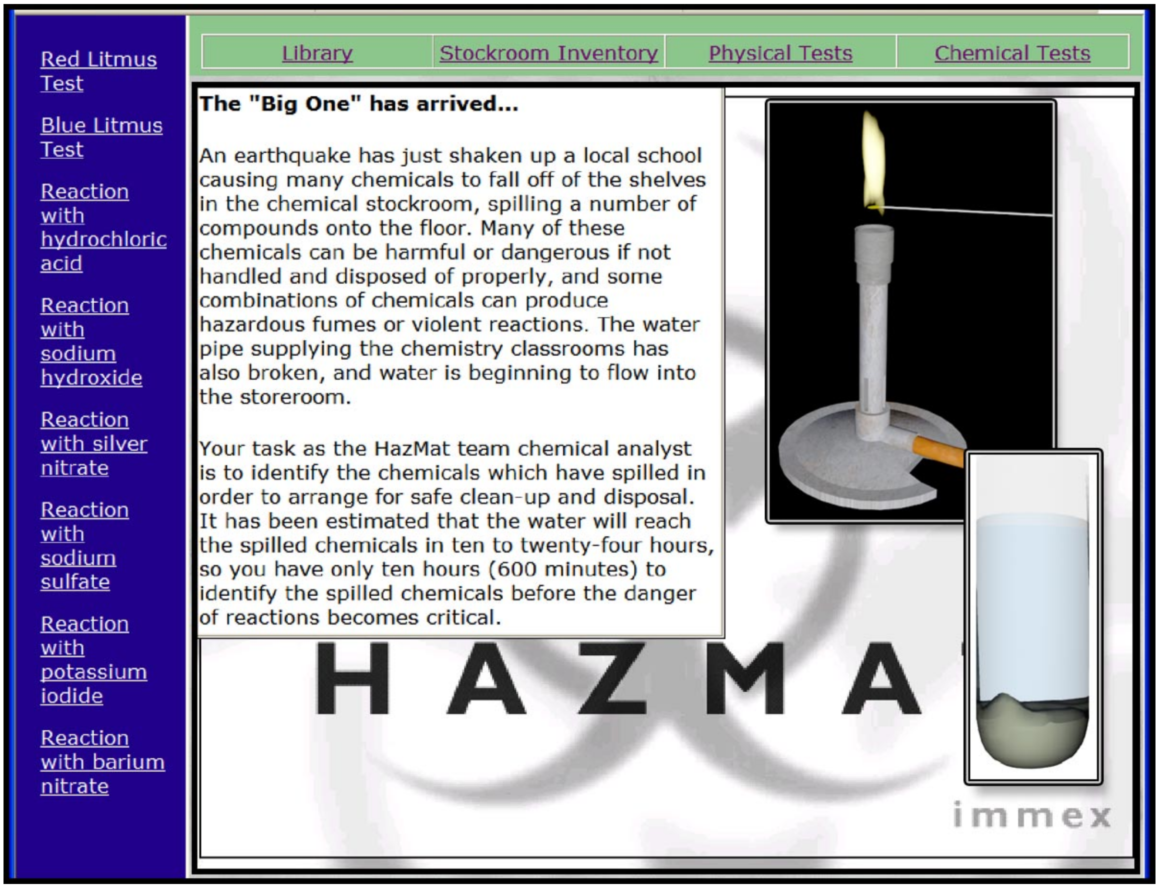

Figure 1. The opening problem-statement screen of the problem used in this study is shown in the upper left-hand corner of the window. This screen states the goal of the exercise to the students and, for this problem set, is identical for all the cases. Also shown are two of the 20 information items available to the students. The item in the upper right-hand corner shows the result of flame testing the unknown and the frame at the lower right of the figure is a precipitation reaction. Although combined into a single window here, students would actually see each item in a separate window while running IMMEX cases and would access each piece of information through the drop-down menus shown at the top of the screen.

As a student proceeds through IMMEX cases, the software's presentation tool records the student's every selection as s/he attempts to solve each case. This feature allows for both real-time and off-line analysis of how students solve a particular case, as well as how student ability changes over time. Since students access IMMEX problems using the World Wide Web, the IMMEX database contains thousands of student performances on hundreds of problem sets in different knowledge domains. While IMMEX records both the informational item a student chose and the order in which s/he chose it, for the purposes of this research a student performance is defined as all the items a student viewed before s/he proposed a final solution to the case being attempted. Information such as the student's answer, date, time and whether the student actually solved the problem correctly is also recorded by IMMEX, in addition to student performance data. The next section will address the tools used to analyze the vast amounts of data recorded by IMMEX.

Methods and Data Sources

In this study, one hundred thirty-four first-year chemistry students at a suburban Southern California high school were asked to identify various unknown chemical compounds using the IMMEX computer simulation software. Student grade point averages, first semester grades, standardized test scores and student demographic data suggest this population of students have characteristics typical of student populations at suburban American high schools with the exception that African American students were under represented and Asian American students were overrepresented in this group (National Center for Education Statistics [NCES], 1999).

The IMMEX problem set used by the students in this study is called *Hazmat* (short for Hazardous Materials). *Hazmat* is a qualitative chemistry problem set in which students are told that there has been an earthquake that has caused a number of chemicals, some of which may be hazardous, to fall off stockroom shelves. As the labels have been obliterated or mixed up with the labels of other compounds and time is of the essence, the school has asked for student help in identifying the spilled chemicals. Each chemical compound represents a unique case in the *Hazmat* problem set. In addition to general stockroom inventory information, there are three physical tests and eight chemical tests the students can conduct on each unknown substance. The students may also review any of eight library reference items in their attempt to identify the unknown. Students must identify the correct unknown from among 57 possibilities on one of two tries at a solution. After the presentation of the problem, students may proceed through the problem space in any manner they choose before ultimately proposing the identity of the unknown. Although the information returned by each menu item may be different from case to case within each problem set, all 23 cases in the *Hazmat* problem set contain the same 20 menu items. In this study, the students' teacher decided to use only those *Hazmat* cases that produced positive results when students chose to conduct a flame test. Both IRT analysis and the teacher's experience suggest these are the easiest of the *Hazmat* cases for this group of students to solve. IRT analysis also suggests these cases are similar, although not identical, to one another in difficulty. Each student received, on average, five individual cases in random order.

While a single student action is occasionally informative in IMMEX problem solving (such as when a student chooses to solve the problem as an initial move and without viewing any information), experience suggests that the sequence of actions or the presence of a group of individual actions are usually much more telling about student understanding (Stevens, Ikeda, Casillas, Palacio-Cayetano, & Clyman, 1999). The number of possible information items in a problem set and the degree of difficulty of the cases a student must solve (as determined by IRT) are generally good indicators of the number of menu items students will choose to view before solving individual cases of a problem set. One average, the students in this study chose to view 17 items before attempting a solution.

Experience also suggests that student performances on cases of IMMEX problem sets are seldom entirely random in nature (Underdahl, Palacio-Cayetano, & Stevens, 2001). In fact, while students may eventually look at all the information contained in a problem space, they will often view menu items sequentially rather than follow a more haphazard strategy. Nevertheless, given the number of pieces of information available to the student, the possible number of unique performances is factorially large. This can make the evaluation of student problem solving difficult. The overwhelming nature of evaluating student problem solving becomes especially apparent when the total number of student performances requiring analysis increases beyond one hundred or so. Consequently, as problem spaces increase in size and student performances multiply, some method must be used to help discern patterns in the data. Cognitive scientists like Fischer and Bidell (1997) suggest that to truly understand such activity, one must develop ways to analyze the patterns of stability and order within the variation of human activity without discarding behavioral complexity. Other literature also suggests that it is unwise to merely compare students or other novices to their teacher or another expert (Glaser & Baxter, 2000).

Novices treat problem solving differently than experts (Messick, 1989; Glaser & Silver, 1994; Baxter & Glaser, 1997), so evaluating novice performances only in terms of how well they match an expert is bound to limit the acceptable student approaches or to discard important complexities of student problem solving. Nevertheless, it is often unclear what features of novice performances will make them successful or effective, and therefore, it is virtually impossible to make estimates about the distributions of these features across performances. This suggests that a non-parametric approach in evaluating such performances is warranted. Therefore, rather than develop a model of behavior and then fit subsequent student performances to that static model, we have chosen to use the demonstrated pattern recognition ability of artificial neural networks (ANN) to identify groups of similar performances in the data (Principe, Euliano, & Lefebvre, 2000).

Our neural network analysis is a two-stage process. First, we use unsupervised neural networks to find clusters of similar performances based on which items of information a student viewed when solving each case. We present a student performance to the ANN as an ordered series of ones and zeros, each representing a menu item. A one indicates the student viewed a particular menu item; zero indicates the student did not. The ANN, in turn, represents this input as a point in 20-dimensional space and finds data clusters by moving a digital marker to the mathematical center of each cluster. The neural net locates this center by moving each marker until the Euclidian distance between a marker and the data point(s) closest to it is minimized. Supervised training is then used to refine the cluster boundaries. Unlike unsupervised training which uses an internal distance metric to determine similarity between performances, in supervised training we present the ANN both a student performance and the cluster that performance is expected to represent. Consequently, we use unsupervised networks as a preprocessor to extract groups of similar performances from the data set, and then feed that information to supervised networks in order to refine the boundaries of the identified clusters. As described elsewhere (Stevens, Lopo, & Wang, 1996; Vendlinski & Stevens, 2000; Vendlinski, 2001), this methodology consistently clusters the same performances together 90% of the time.

Generally, the particular feature(s) of the performances in individual clusters that make them similar are easily discerned. For example, it is common for students to attempt to identify the unknown in a *Hazmat* case with just one test. Thus, a common cluster contains student performances with just this one test. Other clusters are more complex. In order to understand what these clusters represent we used a technique called *mock performance analysis*. In mock performance analysis, we create a performance that represents those features thought to describe each cluster. Then each of these mock performances is fed into the appropriately trained artificial neural network. By adding to and subtracting from each mock performance, the features of each performance that cause the neural net to cluster them together are identified. Using this technique, we have routinely found that 100% of the mock performances cluster where initially anticipated, if the cluster being examined contained at least three performances. Not only does this technique validate the interpretation of each cluster, but, by identifying the salient menu items to describe each cluster, the differences between clusters become readily apparent. In addition, the sensitivity of the clustering network to variations in student performances can be quantified. We term each of the resulting cluster descriptives a *strategy*.

After using artificial neural networks to assign each student performance to a cluster, and ordering these performances chronologically for each student, longitudinal models of student problem solving emerge. In this study, a student's first performance strategy was compared to the most common strategy used by the student to solve all the *Hazmat* cases the student attempted. When viewed individually, this type of analysis addresses the progress individual students make over time. When analyzed as a classroom group, such an analysis describes students more generally, and becomes an indication of class progress or, with multiple classes, teachers and schools, of more generalized learning trends. In these larger student groups, the likelihood students will transition from using one strategy (the beginning state) to subsequently using the same or another strategy is easy to calculate. When represented in condensed form these likelihoods form a transition matrix. Table 4 is an example of such a transition matrix.

By using the transition matrix and elementary matrix multiplication, the distribution of students across clusters can be calculated. With a large enough student sample, both the available literature (Dalphin & Borden, 1997; Soller & Lesgold, 2000) and our prior research suggest it is plausible to make the assumption that without external intervention, the transition likelihoods for a group of students will remain constant in the near term. Consequently, it becomes possible to apply Markov models to determine the distribution of students after each student works a number of successive cases in a problem set. Markov models are used to mathematically represent systems composed of individuals who can be classified into a finite number of states and where the probability that an individual will move between those finite states over time is known. In Markov modeling, the transition probabilities, once determined, are static and transitions are assumed to occur at fixed points within a time interval. The analysis of hundreds of student performances suggests that the probability a student will transition between various states is relatively constant within classes of the same teacher. Therefore, creating a Markov model of such a group is both possible and straightforward. In addition to the transition matrix, the number of students in each state at a particular time can be represented by a distribution vector where each entry in the vector represents the number of students in a particular state at a particular time. The product of multiplying this distribution vector by the complete transition matrix is the distribution vector one can expect after one time interval. In this case, the resulting distribution vector represents the number of students in each state after completing a single *Hazmat* case. Assuming that the transition matrix remains consistent over time, multiplying the resulting distribution vector by the transition matrix again will yield a prediction of the number of students using each strategy after two cases. A third multiplication will yield the distribution vector after three cases, and so on. Although one cannot use this method to trace the path an individual student followed to arrive at a particular strategy type (hence the name hidden chain models), the results of such an analysis is suggestive about the performance of the student group as a whole. Moreover, repeatedly multiplying by the fixed transition matrix generally produces a *steady state*. Exponential multiplication of the transition matrix produces the same result expressed as percentages of students in each state. Consequently, this form allows comparisons of groups of different sizes. Therefore, steady state diagrams provide a metric by which we can evaluate the performances used by a teacher's students. Although not a major focus of this paper, the diagrams would allow comparisons of different classrooms to one another even though different teachers taught the same content to different classes and the students used different strategies to solve the cases in a problem set. Moreover, these diagrams allow us to model not only the effectiveness of the current curriculum, but also the predicted effect of proposed instructional interventions, and how such interventions might ultimately affect the strategies used by these students. This technique is discussed at the end of the next section.

Discussion

Students use a variety of strategies to solve IMMEX cases. Consequently, the student strategies represented by each ANN cluster can vary widely. Nevertheless, many strategies are often closely related and may only differ by one or two items of information. In fact, in some recent performances, the major difference between two strategies was whether a student viewed the problem summary (epilogue). Another major difference between strategies is the success students have solving a case using one strategy or another. For this research, we determined the effectiveness of student strategies by calculating the odds a particular strategy would produce the correct answer to *Hazmat* cases. Good strategies tended to produce correct answers, while poor strategies did not. We use odds here to allow for the comparison of solve rates between different types of cases (as the cases were delivered randomly, not all cases were delivered with the same frequency), and because the natural logarithm of the odds equals student ability (θ) in the one-parameter logistic (Item Response) model. Because the cases in *Hazmat* can be quite dissimilar, students generally will adapt their strategies to account for changes in the cases. For example, a student may be able to identify the unknown compound potassium hydroxide using a flame test and litmus paper, but would have to modify this strategy to identify the compound potassium nitrate. Still, one might reasonably expect that a student who developed the ability to solve one case would demonstrate the ability to effectively solve other, different cases, especially if the student really understood the concepts required to solve these types of problems.

Strategy Types

As expected, in IMMEX problem sets where the cases require students to modify a strategy as the unknown changes, students often do not duplicate a specific strategy exactly; rather the students adapt their strategy to the case they are trying to solve. Nevertheless, there are enough similarities between the different strategies students use to suggest a more general strategy classification scheme might be appropriate. In particular, we have noticed that a number of students use strategies that investigate very few items of information before making an attempt to solve the case. An example of a *limited* strategy in *Hazmat* is the strategy of only viewing the effect of placing the unknown in a flame. While this test can identify the metallic ion in a compound, it tells the student nothing about other parts of the compound. In fact, none of these so called limited types of strategies investigate enough information to conclusively solve the problem. At the other extreme, students investigate more than enough information to solve the problem and often continue to view items even after they have sufficient information to reach a definitive answer. A common example of this strategy is to look at every menu item in the problem space. It is apparent from other research we have conducted (Vendlinski, 2001) that students using this strategy often do not know what information is relevant to hypothesis confirmation and rejection or how to interpret that information. It should also be noted, that although students will sometimes use this strategy to define the boundaries of the problem space in other IMMEX problem sets, we have not seen this behavior in students solving *Hazmat,* possibly because the 20 menu item problem space is relatively small. We have collectively termed these types of strategies *prolific*. Students using either limited or prolific types of strategies are unlikely to solve the case being attempted. In other words, overall students using these strategies had less than even odds of correctly identifying the unknown.

However, when students used strategies that focused only on key pieces of information, they were more likely than not to solve every case they attempted. These strategies were classified as *efficient* types of strategies. Efficient strategies tend to be both case and contextually sensitive. For example, one group of AP chemistry students was very successful in determining the presence of unknown bases by using a strategy which involved the addition of an acid to the unknown, whereas another group of first-year chemistry students at the same school were more successful using a strategy that involved using litmus paper to identify the presence of a base. While both strategies were effective for both groups of students and were very focused, the number of students using each strategy differed between groups. Anecdotally, the AP teacher indicated that when teaching students ways of identifying a base she had stressed the addition of an acid, whereas the teacher of first-year students had focused on the use of litmus paper.

Because strategy types are, in part, determined by how effective they were for the students who actually used them in a particular context, they lend themselves to comparisons across classes as well as comparisons within a single class. While the consequence of this is that a particular strategy could be very effective in one context and ineffective with a different group of students in a different setting, we have not found this to be a common occurrence. In fact, it is more common for certain strategies to be effective no matter where they are used, but to be used by differing percentages of students in various contexts. Our findings are very similar to the cognitive groups Siegler (1998) identified among first grade arithmetic students. The next section illustrates how our findings might be used to develop a formative teaching tool.

Initially, approximately one-third of the students studied here chose to use limited types of strategies, one-third chose prolific strategies, and one-third chose efficient strategies. As expected, a significant number of students who solved the first *Hazmat* case presented to them using an efficient type of strategy, used efficient types of strategies to solve most of the other *Hazmat* cases presented to them. More surprisingly, students using limited or prolific types of strategies predominately continued to use limited or prolific strategy types, respectively, in their attempts to solve future cases *even though those strategies seldom produced a correct answer*. These relationships are shown in Table 1 and, as indicated there, are significant.

Table 1 **Relationship Between First Strategy High School Student Used and the Strategy the Student Used Most Often**

		Most Frequent Strategy		
		Limited	Efficient	Prolific
First Strategy	Limited	30	6	8
	Efficient	5	23	9
	Prolific	5	8	35

Note. This table shows the relationship between the type of strategy a student used to solve his or her first *Hazmat* case and the type of strategy the same student used most often (the mode) to solve subsequent *Hazmat* cases. The distribution around the mode is relatively small. Subsequent to the first performance, on average, 70% of a student's performances were of the same type as the student's modal strategy. Overwhelmingly, without instructional intervention, the strategy type used by a student in her or his initial performance predicts the strategy type that student will continue to use most often on subsequent *Hazmat* cases. The chi-square statistic suggests the relationship is not random. ($\chi^2 = 70.5$; *d.f.* $= 4$; $p < .001$).

This same trend is evident in other classes, at other ability levels and on different IMMEX problem sets. For example, Table 2 shows the relationship between the strategy type first-year undergraduate chemistry students used to solve their first case of a more complex qualitative chemistry IMMEX problem set called *Desperately Seeking Solution* and the strategy type used most often by these same students to solve subsequent cases of this problem set.

Table 2 **Relationship Between First Strategy College Student Used and the Strategy the Student Used Most Often**

		Most Frequent Strategy		
		Limited	Efficient	Prolific
First Strategy	Limited	28	3	7
	Efficient	2	4	1
	Prolific	11	1	28

Note. This table shows the relationship between the type of strategy a first-year undergraduate college chemistry student used to solve his or her first qualitative chemistry case and the type of strategy the same student used most often (the mode) to solve subsequent cases of the same problem set. Overwhelmingly, the strategy type used by a student in her or his initial performance correlates with the strategy type that student continued to use on subsequent cases. ($\chi^2 = 41.96$; *d.f.* $= 4$; $p < .001$).

The degree of difficulty of the cases or problem set (as determined by IRT analysis) also seems to have an effect on the strategy type students use to solve the cases of IMMEX problem sets. Research suggests that as a problem space becomes easier for students to manage (either because the ability level of the students increases or the problem spaces become less complex), students are less likely to use limited types of strategies. In fact when the college undergraduates just described attempted to solve *Hazmat* cases, no student used a limited type of strategy to solve the problem. On the other hand, when the first-year high school chemistry students in this study were asked to solve the more complex qualitative chemistry problem mentioned above, they used almost entirely limited types of strategies to do so. Table 3 shows this:

Table 3 **Relationship Between First Strategy High School Student Used in Solving Complex Case and the Strategy the Student Used Most Often in Solving Complex Case**

		Most Frequent Strategy		
		Limited	Efficient	Prolific
First Strategy	Limited	56	11	2
	Efficient	10	14	3
	Prolific	6	0	12

Note. The same relationship between the type of strategy a high school student used to solve the initial case of the more complex IMMEX problem, Desperately Seeking Solution, and the type of strategy s/he uses most often is evident in this table. As the problem space has become more complex, students begin to favor the use of limited types of strategies. ($\chi^2 = 64.4$; $d.f. = 4$; $p < .001$).

The types of strategies used by a group of students, therefore, seems to suggest whether cases of a particular problem set like *Hazmat* are too easy or too difficult. When almost all students use limited types of strategies to solve presented cases (a type of "floor effect"), the cases or the entire problem set is probably too difficult. Likewise, when most students use efficient or prolific types of strategies to solve cases in a problem set (a type of "ceiling effect"), those cases are probably not challenging enough for the students or indicate that the students have mastered the material. When, however, student performances demonstrate a range of strategy types, one may generally conclude a problem set or a group of cases is appropriate for that student group. Proposed research studies seek to further quantify the association among student ability level, case difficulty, and strategy type selection. Although such considerations may be less important for summative assessment, they are critical when using these results to formulate curricular interventions.

Markov Hidden Chain Analysis

In fact, as will be demonstrated, when the performances of a group of students represent diverse solution strategies, Markov hidden chain analysis could be an effective tool for evaluating the instructional interventions suggested to improve student performances. In this case the three strategy types (limited, efficient, and prolific) and *start* become the states in a Markov model. Start is a holding state for students before they attempt their first IMMEX case. Since the transition probabilities in Markov models show the likelihood an individual will transition from one state to the next, a start state is included so that the model is able to represent how students solve their initial case. Therefore, by definition, students never return to this state. As students attempt other cases in the problem set, they again use one of the three strategy types to solve each new case. The partial state diagram in Figure 2 represents state and transition likelihoods from the class of first-year high school chemistry students in this study.

Figure 2

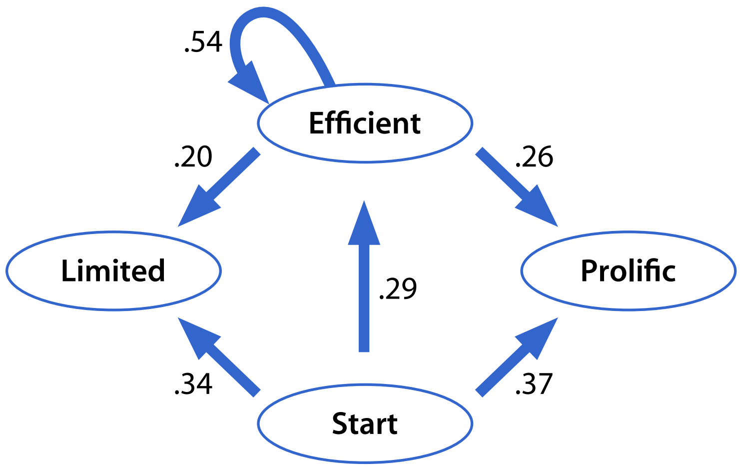

Figure 2. A partial transition state diagram showing the transition likelihoods from two states. The first set of probabilities represent the possibility that students will use a limited (.34), efficient (.29) or prolific (.37) type of strategy to solve their first case of the IMMEX problem, *Hazmat* (i.e., how they transition from the start state). The second set of probabilities represent the likelihood that, once a student has used an efficient type of strategy and so has achieved an efficient state, s/he will use a limited (.20), efficient (.54) or prolific (.26) type of strategy when attempting to solve his or her next case. For the sake of clarity, the transitions from the limited and prolific states are not shown.

The diagram in Figure 2 can be represented by the transition matrix given in Table 4.

Table 4 Probability That Student Will Transition From a Strategy Type

		TO:			
		Start	Limited	Efficient	Prolific
FROM:	Start	**.00**	**.34**	**.29**	**.37**
	Limited	.00	.56	.19	.25
	Efficient	**.00**	**.20**	**.54**	**.26**
	Prolific	.00	.17	.26	.57

Note. The probability that a student in this group will transition from the strategy type given in the left column to the strategy type identified at the top of each column is given where that row and column intersect. The transition probabilities actually shown in Figure 2 are presented here in bold type. As students cannot return to the start state once they solve their first case, the probability that they move from any state back to the start state is 0%. Note also, that for this group, the likelihood a student will use the same type of strategy to solve the very next problem (the diagonal cells of the bottom three rows) is more than 50% for each of the strategy types, suggesting the tenacity of prior strategy types.

To predict and model how the students in this study are distributed among the strategy types after running a single IMMEX case, we multiply the initial distribution vector by the transition matrix. Since all 134 students were initially in the *Start* state, the multiplication can be represented as shown in Equation 1.

(1)

$$\begin{bmatrix}134 & 0 & 0 & 0\end{bmatrix} \bullet \begin{bmatrix} .00 & .34 & .29 & .37 \\ .00 & .56 & .19 & .25 \\ .00 & .20 & .54 & .26 \\ .00 & .17 & .26 & .57 \end{bmatrix} = \begin{bmatrix}0 & 45 & 39 & 50\end{bmatrix}$$

Assuming, as the research suggests, that these probabilities remain consistent over time, the student distribution vector after the seventh case will be $\begin{bmatrix}0 & 45 & 39 & 50\end{bmatrix}$. Multiplying this vector by the transition matrix again will not change this distribution vector. As noted before, this steady state can also be represented by exponential multiplication of the transition matrix. The steady state matrix that results after seven multiplications of the transition matrix by itself is given in Table 5.

Table 5 **Probability That Student Will Transition From a Strategy Type After Seven Cases**

		TO:			
		Start	Limited	Efficient	Prolific
FROM:	Limited	.00	.30	.33	.37
	Efficient	.00	.30	.33	.37
	Prolific	.00	.30	.33	.37

Note. The probability that a student in this group will transition from the strategy type given in the left column to the strategy type identified at the top of each column after attempting to solve seven cases of *Hazmat* is given where that row and column intersect. Assuming that the probability a student will transition from a given state to a subsequent state remains constant during these seven cases (Table 4), the above steady state is reached. The analysis of this group of students suggests that 30% of the students will settle into using limited types of strategies, 33% will use efficient strategies, and 37% will use prolific strategies when solving *Hazmat* cases.

As shown in Table 5, because students tend to adopt a single type of solution strategy from the outset of solving *Hazmat* cases, approximately one third of the students will eventually settle into each of the three strategy types. These steady state transition matrices then become one measure of the current effectiveness of a curriculum. While such steady states would allow for comparisons between classrooms or teachers, a more formative use is also suggested.

Markov Models as Formative Tools

When a teacher analyzes how the strategies identified by artificial neural analysis differ, and combines the insight of that analysis with the Markov technique just demonstrated, various pedagogical interventions can be modeled and the predicted effectiveness of each intervention compared. For example, when reviewing the student performances that generated the data in Table 4, it becomes obvious that one of the efficient strategies students apply to identify part of the unknown is the use of red litmus paper. Red litmus turns blue in the presence of hydroxide, so a successful red litmus test, along with an informative flame test, should allow the student to correctly identify these types of unknowns. Consequently, the teacher may decide to revisit the use and meaning of red litmus in her curriculum. Markov hidden chain analysis allows us to model and predict the outcomes of such an intervention. If we assume, for the moment, that 90% of the students in this class developed an understanding of and could effectively use red litmus after the teacher's intervention, this would imply a change in the types of strategies used by those students when trying to solve *Hazmat* cases that involve hydroxides. More specifically, because the red litmus test would now be meaningful to the students currently using limited or prolific strategies, these students should modify their existing strategy of detecting hydroxide compounds to become more efficient. This change in student behavior would produce the transition matrix in Table 6 and the steady state matrix in Table 7.

Table 6 Probability That Student Will Transition From a Strategy Type Immediately After Instructional Intervention

		TO:			
		Start	**Limited**	**Efficient**	**Prolific**
FROM:	**Start**	.00	.21	.47	.32
	Limited	.00	.39	.41	.20
	Efficient	.00	.24	.49	.27
	Prolific	.00	.12	.44	.44

Note. The probability that a student in this group will transition from the strategy type given in the left column to the strategy type identified at the top of each column from his/her initial to first case (top row), and then from his/her first to second case (bottom three rows) of *Hazmat* after instruction on using red litmus paper is given where that row and column intersect.

Table 7 **Probability That Student Will Transition From a Strategy Type Over Time After Instructional Intervention**

		TO:			
		Start	Limited	Efficient	Prolific
FROM:	Limited	.00	.24	.46	.30
	Efficient	.00	.24	.46	.30
	Prolific	.00	.24	.46	.30

Note. The probability that a student in this group will transition from the strategy type given in the left column to the strategy type identified at the top of each column after instruction on using red litmus paper and after attempting to solve seven cases of *Hazmat* is given where that row and column intersect. Assuming that the probability a student will transition from a given state to a subsequent state (Table 6) remains constant during these seven cases, the above steady state is reached. The analysis of this group of students suggests that almost half the students will use efficient strategies when solving *Hazmat* cases after this specific intervention.

Similar calculations could be made by reducing the percentage of students who benefit from the instruction on using red litmus paper or, if warranted, by applying them differentially to the students using different strategy types. Moreover, the effects of other interventions can be modeled and compared with re-teaching this topic.

Conclusion

Quantitative and qualitative analysis suggest that the strategy types identified by artificial neural network analysis are both accurate and reliable. Moreover, this research suggests that such an analysis could function as a valuable formative tool by allowing us to evaluate teaching interventions designed to benefit both individual students as well as larger, more diverse, groups of students. This study used adaptive artificial neural network analysis to identify the common strategies first-year high school chemistry students used to solve qualitative chemistry problems. It demonstrated that the strategies used by these students were of three general types. Students adopting limited types of strategies did not have enough information to proffer a conclusive answer before doing so. On the other hand, students using prolific strategies had more than enough information to precisely identify the unknown. In both cases, however, students adopting either strategy type were unlikely to correctly identify the compound. Conversely, about one-third of the students in this study adopted very efficient strategies that allowed them to focus only on information that was pertinent to correctly identifying the unknown. Students adopting efficient strategies were more likely than not to identify the unknown substance. These same strategy types are evident in groups of students solving cases from diverse IMMEX problem sets, especially problem sets dealing with content in the science domain.

This study also found that no matter which type of strategy the student used to solve *Hazmat* cases, students would adopt that strategy type beginning with the first case and they would continue to use similar strategies on subsequent cases. This same trend has also been documented in other high school science domains, and among chemistry students of varying abilities (e.g., high school Advanced Placement, community college, and first-year undergraduates). Without instructional intervention, students appear highly unlikely to change problem-solving strategies, *even if those strategies seldom produce a correct answer*. Nevertheless, this research suggests that the analysis of strategies combined with Markov hidden chain analysis could function as a valuable formative tool.

When combined with a teacher's insight of how the strategy types of students differ, Markov analysis not only suggests which interventions might be most effective for students, but also provides a metric that allows us to compare the potential effectiveness of each intervention. Moreover, since we have seen evidence of strategy types in student solutions from the cases of many different problem sets, the same type of analysis should be applicable beyond the quantitative chemistry domain. The methodology proposed here might offer policy makers, investigators in the field, and educators in the classroom, at least in the science classroom, a common metric that allows each to develop and to begin evaluating the effectiveness of current pedagogy, and the effectiveness of proposed interventions.

Note

The authors are grateful to Jennifer Underdahl, Gaelan Lloyd, Stan Chen, John Stallings, Elise Simpson, and Dr. Marcia Sprang for their assistance with this research, and for their help in preparing this manuscript. This research was supported in part by funding from the National Science Foundation (DUE award 0126050) and from the Howard Hughes Medical Institute.

References

American Educational Research Association. (2000). *AERA position statement concerning high-stakes testing in Prek–12 education*. Washington, DC: Author.

Barton, P. E. (1999). *Too much testing of the wrong kind; too little of the right kind in K–12 education*. Princeton, NJ: Educational Testing Service.

Baxter, G. P., & Glaser, R. (1997). *An approach to analyzing the cognitive complexity of science performance assessments*. Los Angeles: CRESST/UCLA.

Bransford, J. D., Brown, A.L., & Cocking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Fischer, K. W. & Bidell, T. (1997). Dynamic development of psychological structures in action and thought. In W. Damon & R. Lerner (Eds.), *The handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 467–561). New York: Wiley.

Glaser, R. & Silver, E. (1994). *Assessment, testing and instruction: Retrospect and prospect*. Los Angeles: UCLA/CRESST.

Lowyck, J. & Poysa J. (2001). Design of collaborative learning environments. *Computers in Human Behavior, 17*, 507–516.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan Publishing Co.

Nathan, M., Koedinger, K., & Martha, W. (2001, April). *The expert blindspot: When content knowledge and pedagogical content knowledge collide*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

National Center for Education Statistics. (1999). Enrollment in public elementary and secondary schools, by race/ethnicity and locale. Retrieved June 2002, from http://nces.ed.gov/surveys/ruraled/data/Race_Ethnicity.asp.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Principe, J. C., Euliano, N.R., & Lefebvre, W.C. (2000). *Neural and adaptive systems*. New York: John Wiley & Sons.

Quellmalz, E., Schank, P., Hinojosa, T., & Padilla, C. (1999). Performance assessment links in science. *Practical Assessment, Research & Evaluation, 6*(10).

Siegler, R. (1998). *Children's thinking*. Upper Saddle River: Prentice Hall.

Stevens, R., Ikeda, J., Casillas, A., Palacio-Cayetano, J., & Clyman, S. (1999). Artificial neural network based performance assessments. *Computers in Human Behavior*, *15*, 295–313.

Stevens, R. H., Lopo, A.C., & Wang, P. (1996). Artificial neural networks can distinguish novice from expert strategies during complex problem solving. *Journal of American Informatics Association*, *3*, 13–38.

Underdahl, J., Palacio-Cayetano, J., & Stevens, R. (2001, April). Practice makes perfect. *ISTE Learning and Leading with Technology*, 26–31.

Vendlinski, T. & Stevens R. (2000). The use of artificial neural nets (ANN) to help evaluate student problem solving strategies. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Proceedings of the International Conference of the Learning Sciences* (pp. 108–104). Mahwah, NJ: Erlbaum.

Vendlinski, T. P. (2001). *Affecting U.S. education through assessment: New tools to discover student understanding*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

About the Authors

Terry Vendlinski and Ron Stevens can be contacted at UCLA/IMMEX Lab, 5601 W. Slauson Avenue #255, Culver City, CA, 90230.

J·T·L·A

The Journal of Technology, Learning, and Assessment

Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College