Assessing Problem-solving Strategies in Chemistry Using the IMMEX System

Melanie Cooper - *Clemson University*
Ron Stevens - *University of California, Los Angeles*
Thomas Holme* - *University of Wisconsin, Milwaukee*

ABSTRACT

This paper outlines the development and utility of a problem-solving assessment system called IMMEX™ (Interactive MultiMedia Exercises). This system presents students with a complex problem to solve in an online environment. Student actions are tracked and data-mining strategies allow the clustering of many possible problem-solving pathways into specific strategy types. The system provides reliable and repeatable measures of student problem solving, which can be used to determine effective teaching strategies or to evaluate research studies in chemistry. Future developments that will allow the system to be part of a comprehensive assessment strategy within an undergraduate chemistry curriculum are also possible.

THE CHALLENGE

All stakeholders in science education efforts recognize the need for developing effective problem solvers; yet instructors continue to find it challenging to quantify students' strategic thinking in ways that can rapidly inform, and help modify instruction [1]. The current challenges for such assessments relate to the cognitive and non-cognitive differences among students, the difficulty of generalizing across problem-solving content domains, the design and development of appropriate tasks, and finally the speed, scale, and conceptual accessibility of assessment data.

Strategic problem solving is influenced by many variables, such as: students' prior knowledge and skill, cognitive and metacognitive abilities, task characteristics, gender, ethnicity, and classroom environment [2], as well as affective variables such as motivation and self efficacy [3]. An additional complication is that the acquisition of problem-solving skills is a dynamic process characterized by transitional changes over time as experience is gained and learning occurs [4].

While the challenges for developing assessments of problem solving are substantial, the real-time generation and reporting of metrics of problem-solving efficiency and effectiveness could fulfill many of the purposes for which educational assessments are used, such as grading and feedback for improving learning. [5,6]. In addition, such metrics could help target interventions for students, focus professional development activities for teachers, and influence training, implementation, and support decisions throughout school districts [7].

Effective and efficient problem solving in a domain requires both declarative (the 'what is') knowledge of the domain, as well as the ability to formulate and execute strategies (the procedural knowledge or the 'how to') to accomplish the goals posed by different problems. As expertise is developed there is a continual interplay between these two knowledge forms resulting in the development of refined representations or schema, which are defined as structures that allow problem solvers to recognize a problem state as belonging to a particular category of problem states that normally require particular moves. Such schema can serve as useful guides or roadmaps for future problem-solving situations.

In most undergraduate science courses the 'what is' is acquired through classroom instruction and outside readings, and with modern psychometric techniques, students' declarative abilities can be measured with some precision and reliability. Procedural knowledge is best gained through practice where elements of declarative knowledge can be combined and manipulated in various ways to solve common or not so common problems within the domain. In undergraduate courses these skills are often practiced in laboratory situations ranging from directed inquiry to complete inquiry. Procedural knowledge can be more difficult to assess than declarative knowledge, because determining if and how students failed at a complex task is equally as important as how they succeed.

THE IMMEX PROJECT

The IMMEX Project began in 1988 as a way to test the emerging diagnostic skills of medical students in ways other than Multiple Choice Testing, that is, as a way of promoting and assessing procedural skills. IMMEX problem solving follows the hypothetical-deductive learning model of scientific inquiry [8] where students 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.

The first IMMEX simulations presented patient diagnosis scenarios in cellular and molecular immunodeficiency [9] and were programmed in Turbo Pascal. These initial simulations had full database capabilities that could be queried by Structured Query Language (SQL) to provide a glimpse of not only if the student solved the problems, but more importantly, how they solved the problems [10].

The subsequent intellectual evolution of IMMEX reflects activities occurring along multiple parallel dimensions or strands relating to 1) grade and discipline expansion of the problem-solving concept, 2) the refinement of the visualization and data mining tools to better understand the idea of student problem-solving performance and progress, and 3) scale up of all aspects of IMMEX. The current ASA grant benefits from progress made by the IMMEX Project along all three dimensions.

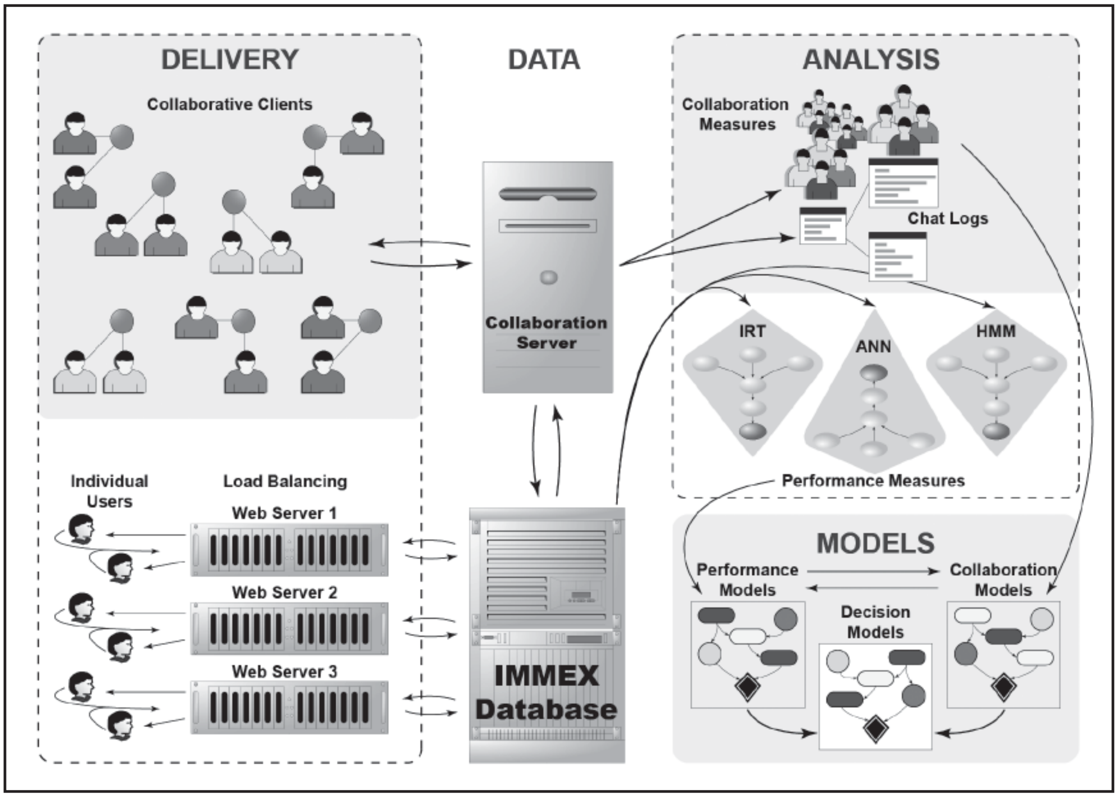
Strand 1: Development and Implementation of IMMEX Simulations

The first strand addresses the expansion of the problem-solving program from a medical school environment to one covering the K-16 educational continuum. Events in this strand include adaptations of the interface and presentation modes to accommodate the needs of different audiences as well as different forms of professional development activities to make the problems and data accessible to a wide range of audiences (students, teachers, parents, administrators). Activities in this strand also relate to the specific types of funding needed to support these events.

From a systems architecture perspective, IMMEX is a data-centric system centered around an SQL database containing problem components and performance data. It consists of Delivery, Data, Analysis, and Modeling components, connected with direct or web services communications. In the delivery component, options are currently available for 1) large scale (~400 concurrent users) individual problem solving and 2) for pilot testing (~20 concurrent student groups) for collaborative problem solving.

The IMMEX database server records the student performance data and the collaboration server records the student chat log. These are subsequently merged during the chat modelling process to associate chat segments with the test selections. For collaborative studies, the collaboration client runs in a browser and is managed through Java applets that communicate with the IMMEX Collaboration Server. The Collaboration Server is an HTTP server acting as a proxy, that filters, edits, and synchronizes the IMMEX HTML pages through JavaScript, and sends them to the clients.

Figure 1. Overall Architecture for IMMEX delivery and Assessment Modules



The analytic models that provide the engine for suggesting interventions, focus on 1) effectiveness, as measured by Item Response Theory (IRT) analysis, and 2) strategies, as modeled by Artificial Neural Network (ANN) analysis and Hidden Markov Modeling (HMM). We have chosen to model both in real time, but in different software modules, as we think they may be assessing different constructs [11]. The analyzed data can then be propagated and integrated back into the decision models as described below, for providing or triggering interventions as needed.

Strand 2: Development of a Layered, Multidimensional Assessment System

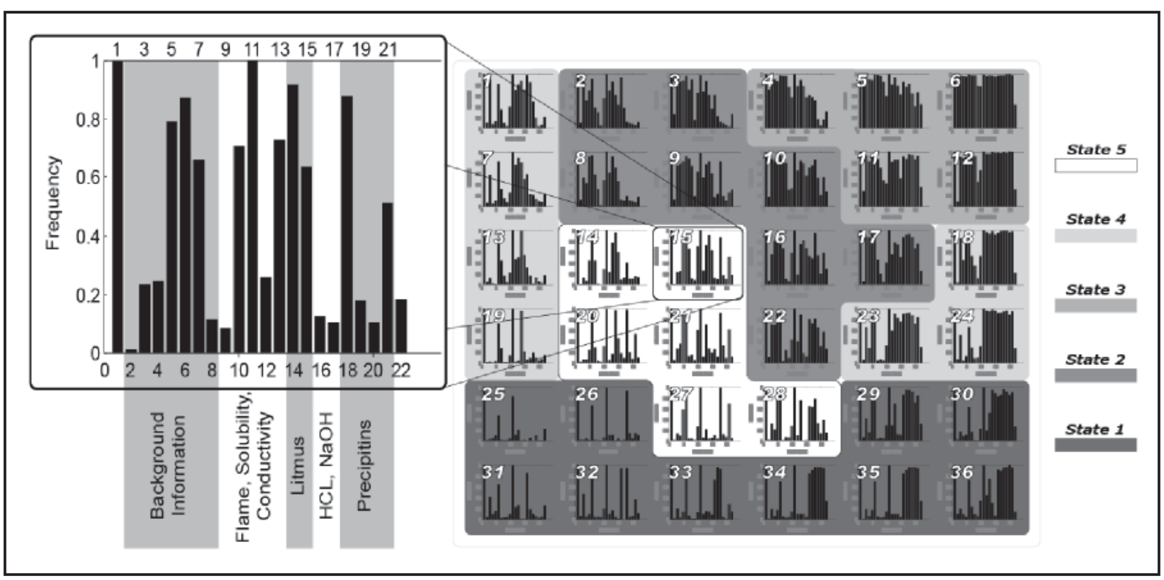
The second strand traces the evolution of a layered system of data analysis and visualization techniques that facilitates the extraction of information from the ever-expanding dataset. We believe that the paths that students employ while navigating an IMMEX task provide evidence of a strategy, which we define as a sequence of steps needed to identify, interpret, and use appropriate and necessary facts to reach a logical conclusion or to eliminate or discount other reasonable conclusions. From these strategies a student demonstrates understanding by consistently, and efficiently deriving logical problem solutions.

In addition to providing students with opportunities to engage in scientific problem solving, IMMEX supports detailed assessments of students' problem-solving skills and learning trajectories [12,13,14]. Although there are dozens of ways to approach and solve IMMEX problems, the cases are closed ended or well defined in that a problem is either solved or not solved. Each IMMEX "problem set" contains 6 to 50 "cases" of the problem with the same interface and resources, but with different unknowns and supporting data [15]. IMMEX includes a layered system of analytic tools to dynamically model various aspects of students' problem-solving performance, including: 1) The strategic sophistication of a student at a particular point in time (a performance measure), 2) How the student arrived at this level (a progress measure), 3) How s/he will likely progress with more practice/experience (a predictive measure), 4) Whether this strategic level will be maintained (a stabilization measure), and 5) What instructional interventions could most effectively accelerate each student's learning. And then, of course, comes the next challenge of how to generalize such assessments across domains and educational systems [14].

IMMEX modeling utilizes serialized timestamps of how students use the resources that are available in an individual problem set, and across the problem's cases. For each problem, models are formed in real time based on 1) estimates of student ability, 2) the strategies used, and 3) estimates of future performance. These metrics are generated through IRT analysis, ANN analysis, and HMM, respectively [12,13,14]. For IRT analysis the problem difficulties are first estimated for all cases using the solve rates from a large number of student performances. Then, using this model, the ability of each student is estimated based not only on whether or not the case was solved, but also the relative difficulty of the case.

As students solve IMMEX cases, the menu items selected are also used to train competitive, self-organizing ANN analyses [16,17]. Self-organizing maps learn to cluster similar performances in such a way that neurons near each other in the neuron layer respond to similar input vectors [18]. The result is a topological arrangement of performance clusters where geometric distance between these clusters becomes a metaphor for strategic similarity. We frequently use a 36-node neural network and train with between 5,000 and 10,000 performances derived from students with different ability levels and where each student performed at least 3-4 cases of the problem set [19]. The components of each strategy in this classification can then 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 for Identifying Strategies. The selection frequency of each action (identified by the labels) is plotted for the performances at node 15, thus characterizing the performances for this node and relating them to performances at neighboring nodes. The nodes are numbered in rows, 1-6, 7-12, etc. This figure also shows the item selection frequencies for all 36 nodes [12,13,14].



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, many Node 15 performances shown in Figure 2 (left side) contain the items13-15, whereas few contain items 16 and 17 or 19 and 20. Also shown 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 IMMEX problem sets contain many parallel cases, learning trajectories can be developed through HMM that not only reflect and model students' strategy shifts as they attempt series of cases, but also predict future problem-solving performance [12].

Prior analyses indicate that many students begin by selecting many test items as they attempt to solve IMMEX problems. Consistent with models of skill acquisition [20], over time most students refine their strategies and select fewer tests, and eventually stabilize with a preferred approach often used on subsequent problems. As expected, with practice, students' solve rates increased from 35 percent to 63 percent ($\chi 2 = 121.8$, df=10, $p<0.000$). The rate of stabilization, and the strategies stabilized with are influenced by gender [21], experience [12], and individual or group collaboration [22], etc. Students often continue to use these stabilized strategies for prolonged periods of time (3-4 months) when serially re-tested [14]. Significant variability of strategy usage is also seen across teachers and classrooms, and detailed videotape analysis of instruction suggests that the ways teachers represent the task to the students has a major effect on the problem-solving outcomes [23].

Strand 3: Scaling Up and Scaling Out

The third strand involves scale and intersects in different ways with the other two strands. Part of the current ASA effort relates to this Strand.

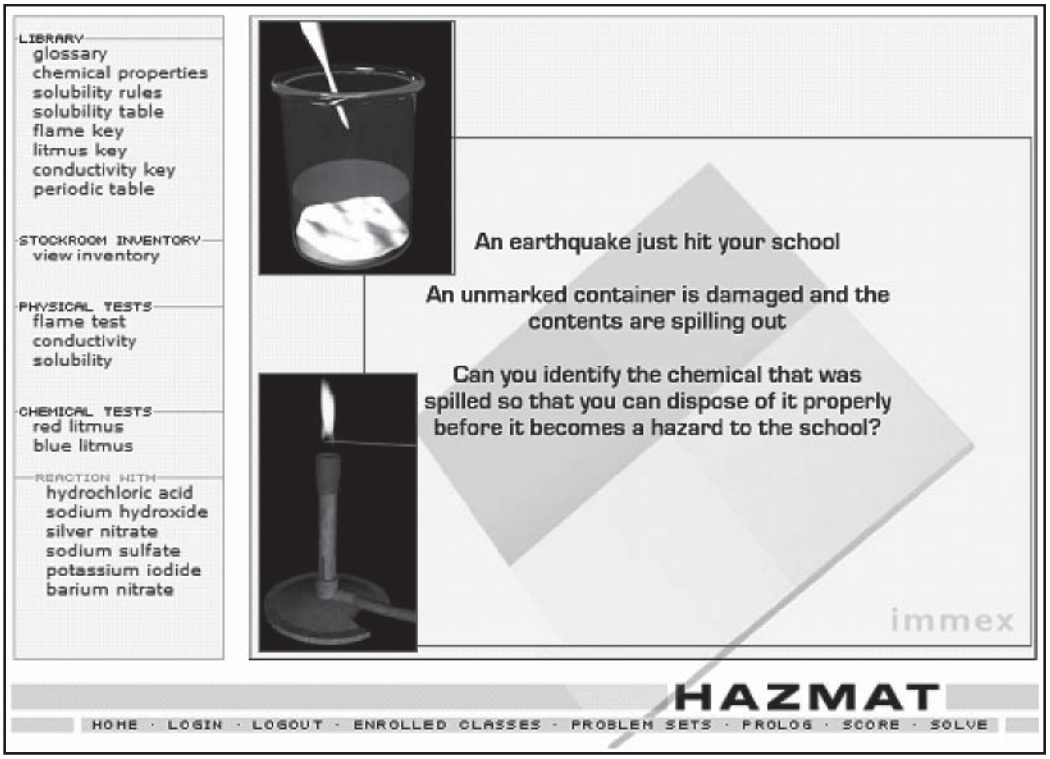
Strand 4: Implications for Teaching and Learning

The development of IMMEX problems and the suite of assessment tools that accompany them has profound implications for both the summative and formative assessment of student problem solving.

SUMMATIVE ASSESSMENT

We now have over thirty IMMEX problem sets covering topics in general and organic chemistry (see http://chemed.ces.clemson.edu/chemed/IMMEXWorkbook.pdf for a workbook detailing the problems in these subjects. There are also many other problems in other fields.) and it is now possible to rapidly assess student performance on complex open-ended problems. An example of a typical IMMEX problem is given in Figure 3. This problem, named Hazmat, is a qualitative analysis problem, in which students must choose appropriate tests to identify an unknown. The results of the tests are presented as animations, so that the student has to understand what the results mean, and act accordingly.

Figure 3. The prolog for Hazmat, a qualitative analysis problem.



The ability to assess complex problem-solving abilities should have a major impact on teaching and learning, because it is a truism that students will learn in response to how we assess them, or to put it another way, assessment often drives the curriculum. If we assess students with multiple-choice problems that emphasize recall and simple one-step problem solving, students will not learn the skills which are so valued in scientists; that is synthesis evaluation and critical thinking. Most online assessment systems grade only the answer to the problem, and even though many now allow free-form input of answers, most systems are still very limited in the complexity of the problem that can be posed. In fact, the systems that pose a different problem set to each student actually encourage algorithmic problem-solving behaviors, since all that changes from problem to problem is the numbers. We now have available to us a system that will allow educators to assess students' complex problem-solving abilities and strategies, while monitoring the strategy that the student chooses to navigate through the problem. The output is seamless and can be used to assess the achievement of the student in a particular topic or course.

FORMATIVE ASSESSMENT

The ongoing assessment and feedback capabilities of IMMEX are probably even more important than the end-of-course type of traditional exam or test. IMMEX provides us with tools to probe how a student solves complex problems, whether that student changes strategy over time, and whether the student actually improves in ability over time. It can give us insight into how students solve problems and it can tell us whether an intervention or teaching strategy is effective in producing improvements in ability and strategy choice. We can measure the effects of interventions and have direct evidence that student problem-solving behaviors can be affected by relatively minor interventions. We can use IMMEX to measure improvements in problem solving and provide feedback and support to students who are not progressing well.

RESEARCH RESULTS

- *Students stabilize at a particular ability level and strategy after about five attempts at a problem*

We have found that a student will often stabilize on a problem-solving strategy (as modeled by ANN and HMMs, [12,24,25]) and at a particular ability level (from IRT, [22,26]) by the time she has performed five problems from that particular problem set. That is, even if the student is not improving, and is using an inefficient and/or ineffectual problem-solving strategy, that student will rarely change strategy once they have stabilized. Nor will the student improve in ability. This finding seems to belie the commonly held belief that if only the student would work harder or longer, then they would improve.

- *Student problem-solving ability and strategy choice can be changed permanently by interventions such as collaborative learning*

One area of our research that has profound implications for teaching and learning is the investigation of interventions to determine whether we can perturb the stabilization levels of problem-solving strategy and ability. For example, we investigated the effect of collaborative grouping on students' problem-solving strategies and abilities [22,25,26]. This research was performed with 713 students enrolled in the first semester of a general chemistry course for science and engineering majors.The experimental design was a modified pre-test, intervention, post-test design, in which students were asked to perform five or more IMMEX problems to provide a baseline stabilization of strategy and ability. This was followed by the intervention in which students worked in pairs on five more problems, and then the "post-test" in which students again worked five or more problems individually. In this way we were able to identify any changes that took place as a result of grouping, and whether these changes remained after grouping.

Figure 4. Strategies used by individuals who have stabilized, before and after grouping. Key shows strategies 1-5 resulting from ANN analysis followed by HMM of the output. The strategies are significantly different (Chi square = 227, p<0.001)

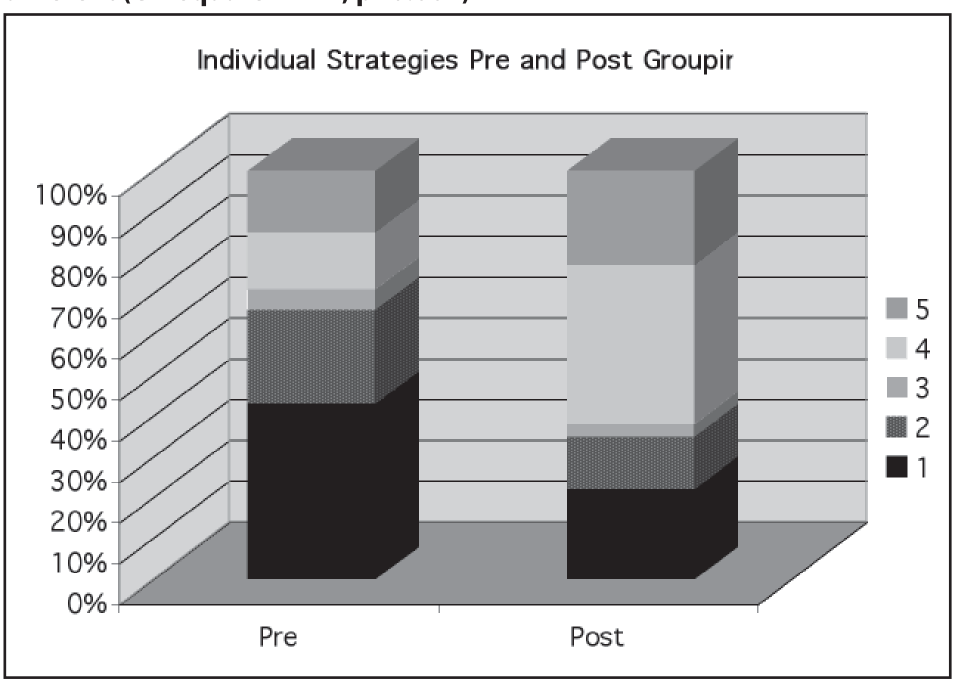
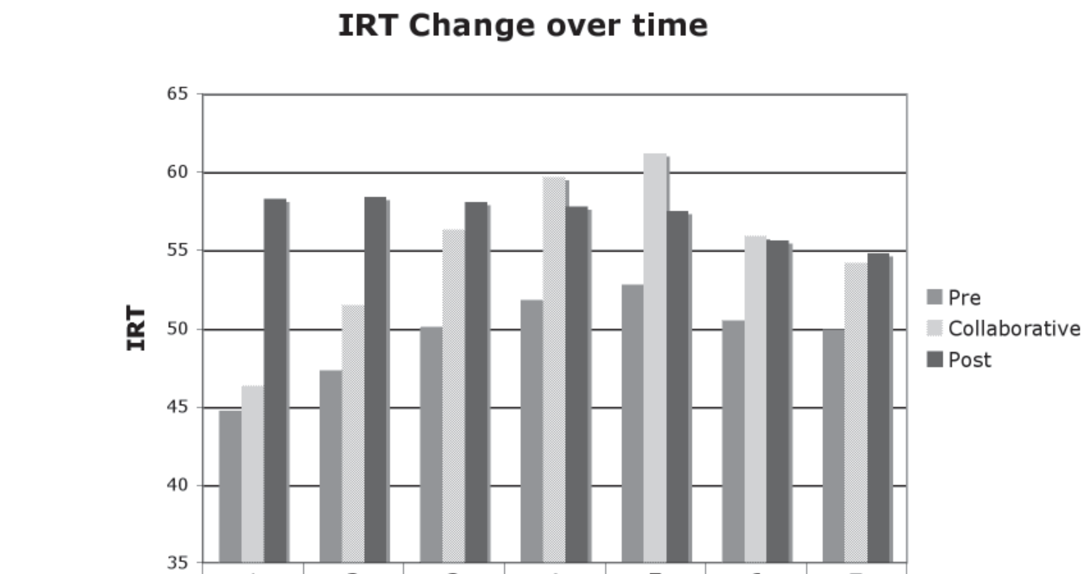


Figure 5. A comparison of student abilities for pre-collaborative, collaborative, and post-collaborative performances. The table shows the actual abilities from IRT analysis, which are reported on a scale ranging from 20-80.



	1	2	3	4	5	6	7
Pre	44.7	47.3	50.1	51.8	52.8	50.5	49.9
Collaborative	46.3	51.5	56.3	59.7	61.2	55.9	54.2
Post	58.3	58.4	58.1	57.8	57.5	55.6	54.8

Problem number

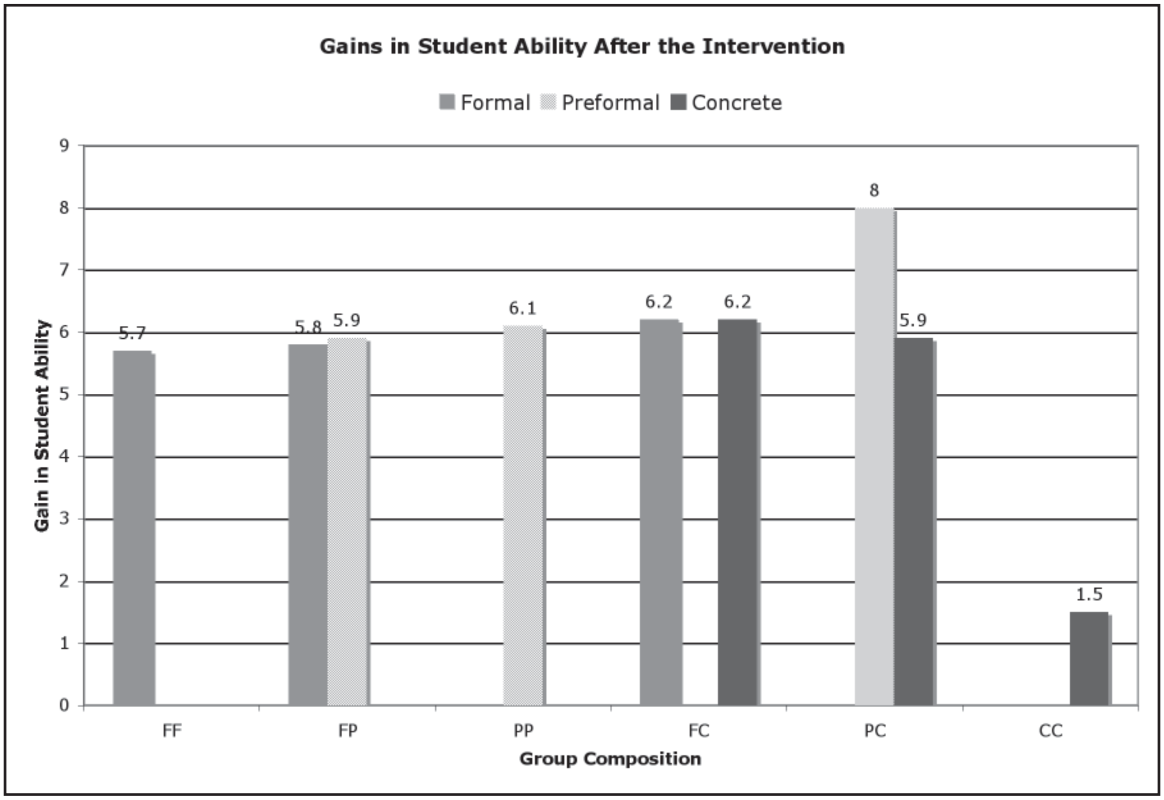
Although the results reported here are from a study involving 713 students, we have now performed this type of experiment with literally thousands of students, and the results are always that students tend to change strategy and increase in ability during grouping, and that the strategy changes and ability improvement remain after the intervention, as shown in Figures 4 and 5.

Thus, one intervention consisting of a one- to two-hour problem-solving session with an unstructured group, had the effect of improving the students' ability by about 6 IRT ability units, or about 10 percent. This is direct evidence that an unstructured group problem-solving session can be beneficial. The results indicate that for most students, some part of their time is much better spent solving problems in a group setting rather than plugging away at a problem set alone. This is not to say that group work can substitute for individual effort, but clearly it is beneficial for most students.

- *The student and group makeup can affect improvements*

Another interesting and important finding from this research was that student improvement depends upon the makeup of the group. In these experiments students were paired with another student by their level of logical thinking, as determined by the GALT test. Students were classified as Formal, Transitional, or Concrete, and paired up in all possible combinations. Remarkably, the makeup of the group did not seem to affect the improvement much, except for two types of combination. For most students the improvement is about 10 percent, or 5-6 IRT ability units, as can be seen from Figure 6, which shows the improvement for students in IRT ability units (which generally range from 20-80 with an average in our studies of 50 units). However, if a transitional student was paired with a concrete student, the improvement was much greater, while if two concrete students were paired there was no significant improvement at all.

Figure 6. Gain in student ability for each type of group and the thinking level of the students in the group. The gain is statistically significant for every group at the p<0.001 level, except for the C-C group.



PROGRAM ASSESSMENT AND THE FUTURE OF IMMEX IN CHEMISTRY

A growing number of educational reports address the need to consider assessment within higher education. The recent report from the Spellings Commission on Higher Education [27], for example, calls for improved accountability within higher education, among several recommendations. As in many disciplines, chemistry departments are searching for ways to implement enhanced assessment prerogatives. In addition to pressure arising from accreditation and national educational policy directives, the process by which the American Chemical Society approves undergraduate chemistry curricula for students is undergoing change as well [28]. This development also places additional impetus on chemistry departments to establish the efficacy of their educational efforts for outcomes-based reporting.

The American Chemical Society, through its Division of Chemical Education, also provides the chemistry education community with assessment materials on a national scale via the Examinations Institute, currently housed at the University of Wisconsin – Milwaukee. The Exams Institute produces exams for most undergraduate chemistry courses through a committee writing process [29]. These exams, however, tend to present snapshots at the end of a course, rather than a measure across sub-disciplinary boundaries of student growth.

Thus, one envisioned pathway for the expansion of IMMEX is to provide a means for chemistry departments to measure student problem-solving strategies in several courses of the undergraduate curriculum. Comparison of how those strategies change with increased exposure to content knowledge and skills development associated with the undergraduate curriculum would provide a measure of growth. Insofar as the strategy measures made possible by IMMEX assess generalized knowledge (not constrained by the sub-discipline of chemistry, in which the problem is set) the use of IMMEX potentially provides a powerful tool for measuring growth of knowledge. Efforts are currently underway to establish the requisite concepts of skill-transfer in a specific category of chemistry problems – those involving the relationship between structure and function of molecular systems. Transfer is a cognitive issue with significant complications [30,31], but the idea of measuring strategies for students as they progress in the chemistry major is straightforward in the context of IMMEX. Thus, as more IMMEX problems are staged for later courses in chemistry, the prospect that meaningful metrics for measuring growth in the sophistication of problem-solving strategies remains.

SUMMARY

This paper has shown that within a complex, online environment where student actions can be followed, it is possible to derive a large database of student performances on a class of chemistry problems and subsequently use data-mining methodologies to establish general strategies that are represented among all the thousands of pathways for solving the problem. The IMMEX system that was devised to carry out this type of measurement has established that valid and reliable measures of problem-solving strategies can be made within this environment. Chemistry education research has shown that the IMMEX system can make measures of student strategies and then be used to determine if teaching interventions are helpful in improving those strategies. This methodology shows promise for future expansion to allow the IMMEX system to be incorporated as part of an overall assessment plan for an academic chemistry department that wishes to measure the change in student problem-solving skills over time.

ACKNOWLEGMENTS

This project has been funded by several grants from the National Science Foundation (DUE- 01236050, REC-02131995, NSF-ROLE 0528840, HRD-0429156, DUE-0512526), whose support is gratefully acknowledged.

REFERENCES

[1] Pelligrino, J.W., Chudowsky, N., and Glaser, R., *Knowing what students know: The science and design of educational assessment*, (National Academy Press, Washington, D.C.), 2001.

[2] Fennema, E., Carpenter, T.P, and Jacobs, V.R. , "A longitudinal study of gender differences in young children's mathematical thinking," *Educational Researcher*, vol. 27, pp. 6-11, 1998.

[3] Mayer, R.E., "Cognitive, metacognitive, and motivational aspects of problem solving," *Instructional Science*, vol. 26, pp. 49-63, 1998.

[4] Lajoie, S.P., "Transitions and trajectories for studies of expertise," *Educational Researcher,* vol. 32, pp. 21-25, 2003.

[5] Bennett, R.E., "Reinventing Assessment. Speculations on the Future of Large-Scale Educational Testing. A Policy Information Perspective," *Educational Testing Services*, 1998, Available on ERIC, ED424254.

[6] Atkin, J.M, Black, P., and Coffey, J., *Classroom Assessment and the National Science Education Standards,* (National Science Foundation, Arlington, Va.), 2001.

[7] Spillane, J.P., Reiser, B.J., and Reimer, T., "Policy implementation and cognition: Reframing and refocusing implementation research, " *Review of Educational Research*, vol. 72, pp. 387-431, 2002.

[8] Lawson, A.E., *Science Teaching and the Development of Thinking,* (Wadsworth Publishing, Belmont, Calif.), 1995.

[9] Stevens, R.H., Kwak, A.R., and McCoy, J.M., "Evaluating preclinical medical students by using computer-based problem-solving examinations,"*Academic Medicine*, vol. 64, pp. 685-687, 1989.

[10] Stevens, R.H., McCoy, J.M., and Kwak, A.R., "Solving the problem of how medical students solve problems," *MD Computing*, vol. 8, pp. 13-20, 1991.

[11] Stevens, R.H. and Thadani, V., "A Bayesian network approach for modeling the influence of contextual variables on scientific problem solving," Lecture Notes in Computer Science, vol. 4053, pp. 72-84, 2006.

[12] Stevens, R., Soller, A., Cooper, M., and Sprang, M., "Modeling the development of problem-solving skills in chemistry with a web-based tutor," *Lecture Notes in Computer Science*, vol. 3220, pp. 580-591, 2004.

[13] Stevens, R.H., Johnson, D.F., Soller, A., "Probabilities and Predictions: Modeling the Development of Scientific Competence," *Cell Biology Educ*., vol. 4, pp. 42 – 57, 2005.

[14] Stevens, R.H., "Constructing Machine Learning Systems for Quantifying Student's Scientific Problem-solving Efficiency and Effectiveness," *Technology in Cognition and Instruction,* in press.

[15] Stevens, R., Palacio-Cayetano, J., "Design and performance frameworks for constructing problem-solving simulations," *Cell Biology Education*, vol. 2, pp.162–179, fall 2003.

[16] Stevens. R., and Najafi, K., "Artificial Neural Networks as Adjuncts for Assessing Medical Students' Problem-solving Performances on Computer-based Simulations," *Computers and Biomedical Research,* vol. 26, no. 2, pp. 172-187, 1993.

[17] Stevens, R., Wang, P., Lopo, A., "Artificial neural networks can distinguish novice and expert strategies during complex problem solving," *JAMIA*, vol. 3, no. 2, pp. 131-138, 1996.

[18] Kohonen, T., *Self Organizing Maps*, 3rd extended ed., (springer, New York), 2001.

[19] Stevens, R., and Casillas, A., "Artificial Neural Networks," in Mislevy, R.E., Williamson, D.M., and Bejar, I. (eds.), *Automated Scoring*, (Lawrence Erlbaum, Mahwah, N.J.), 2006.

[20] Ericsson, K.A., "Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains," *Academic Medicine*, vol. 79, pp. S70-S81, 2004.

[21] Stevens, R.H., and Soller, A., "Machine learning models of problem space navigation: The Influence of Gender," *ComSIS*, vol. 2, no. 2, pp. 83-98, 2005.

[22] Cooper, M.M., Cox, C.T., Nammouz, M., Case, E., and Stevens, R., "An Assessment of the Effect of Collaborative Groups on Students' Problem-solving Strategies and Abilities," *J. Chem Education*, submitted.

[23] Thadani, V., and Stevens, R.H., "Teaching Practices and Their Relationship to Students' Scientific Problem-solving Skills", submitted.

[24] Case, E.L. Cooper, M.M. Stevens, R.H., "Is Collaborative Grouping an Effective Instructional Strategy," *J. Coll. Sci. Teach.*, vol. 36, pp. 42-47, 2007.

[25] Nammouz, M. Ph.D. Dissertation, Clemson University, 2005

[26] Cox, C. T. Ph.D. Dissertation, Clemson University, 2006

[27] U.S. Department of Education, *A test of leadership: Charting the future of US higher education*, 2006. Available at http://www.ed.gov/about/bdscomm/list/hiedfuture/index.html

[28] American Chemical Society, "Proposed Revision to the ACS Guidelines," 2006, available online at http://www.chemistry.org/portal/a/c/s/1/acsdisplay.html?DOC=education\cpt\Proposed_Revision_to_the_ACS_Guidelines.html

[29] Holme, T.A., "Assessment and Quality Control in Chemistry Education," *J. Chemical Education*, vol. 80, pp. 594-598, 2003.

[30] Catrambone, R., "Generalizing Solution Procedures Learned From Examples," *J. Experimental Psychology,* vol. 22, pp. 1020-1031, 1996.

[31] Sohn, Y.W., Doane, S.M., and Garrison, T., "The impact of individual differences and learning contexts on strategic skill acquisition and transfer," *Learning and Individual Differences*, vol. 16, pp. 13-30, 2006.