

Articles

Design and Performance Frameworks for Constructing Problem-Solving Simulations

Ron Stevens* and Joycelin Palacio-Cayetano

The IMMEX Project, UCLA School of Medicine, UCLA Graduate School of Education, 5601 West Slauson Avenue, No. 255, Culver City, California 90230

Submitted February 3, 2003; Revised March 12, 2003; Accepted March 13, 2003
Monitoring Editor: Eric H. Chudler

Rapid advancements in hardware, software, and connectivity are helping to shorten the times needed to develop computer simulations for science education. These advancements, however, have not been accompanied by corresponding theories of how best to design and use these technologies for teaching, learning, and testing. Such design frameworks ideally would be guided less by the strengths/limitations of the presentation media and more by cognitive analyses detailing the goals of the tasks, the needs and abilities of students, and the resulting decision outcomes needed by different audiences. This article describes a problem-solving environment and associated theoretical framework for investigating how students select and use strategies as they solve complex science problems. A framework is first described for designing on-line problem spaces that highlights issues of content, scale, cognitive complexity, and constraints. While this framework was originally designed for medical education, it has proven robust and has been successfully applied to learning environments from elementary school through medical school. Next, a similar framework is detailed for collecting student performance and progress data that can provide evidence of students' strategic thinking and that could potentially be used to accelerate student progress. Finally, experimental validation data are presented that link strategy selection and use with other metrics of scientific reasoning and student achievement.

Keywords: cognitive frameworks, problem solving, science education, undergraduate, strategy.

INTRODUCTION

The knowledge needed to solve problems in a complex domain such as biology or chemistry is composed of many principles, examples, technical details, generalizations, heuristics, and other pieces of relevant information. From a cognitive perspective, these components can be broadly grouped into factual (declarative), reasoning (procedural), and regulatory (metacognitive) knowledge/skills (Anderson, 1980), and all play complementary roles in the construction of such science knowledge. Declarative knowledge is characterized by what people can report (knowing that) and facilitates the construction of organized frameworks of science concepts while providing scaffolding for the acquisition of new concepts (Novak and Gowin, 1984). Teacher-centered instructional practices are often effective at developing declarative knowl-

edge and benefit students who require a more structured-learning approach (Von Secker and Lissitz, 1999).

Procedural knowledge is characterized by knowledge that appears in a performance but cannot be easily reported (knowing how) and, along with regulatory skills, manifests itself as strategy selection and utilization mechanisms associated with hypothesis-driven (Lawson *et al.*, 2000) and goal-oriented situations. In scientific reasoning, the declarative and procedural knowledge exists as a continuum from domain general skills (for instance, how to graph results) to domain specific components (how to analyze flow cytometry graphs). These continua help benchmark the relative levels of professional expertise that different users apply to complex simulations.

These reasoning skills are important for achievement in science (Staver and Halsted, 1985; Krijik and Haney, 1987) and students may require extended experience for maximum observable effects. For instance, Johnson and Lawson (1999) have shown that reasoning abilities accounted for more

DOI: 10.1187/cbe.03-02-0006

*Corresponding author. E-mail address: immex_ron@hotmail.com.

achievement variance in a community college biology course than did either prior knowledge or the number of biology courses taken. Improvement in reasoning skills has also been shown to occur as a result of prolonged instruction and can lead to long-term gains in science achievement (which may exhibit transfer to other disciplines) (Shayer and Adey, 1993). This indicates that the duration and intensity of exposure to reasoning situations are important factors for the development of reasoning skills and that more individually targeted interventions may enrich/personalize the process. This suggests the need to provide students with diverse, continual and prolonged problem-solving experiences; others (Pellegrino *et al.*, 2001) would also argue the need to begin routinely assessing students in such formats.

Instructional practices for developing scientific reasoning skills take many forms, including laboratory, inquiry-based science, and, increasingly, computer simulations, but generally share the properties of being student-centered, socially interactive (Seely-Brown *et al.*, 1989), and constructivist, involving many opportunities to gather evidence to support claims, analyze data quantitatively, and construct explanations. Problem-solving approaches provide such rich reasoning environments and have been used as effective instructional strategies in professional schools for some time (Elstein, 1993; Kolodner, 1993). These approaches have been attractive and effective not only for the motivational and realistic contexts that they provide but also for the balance of knowledge styles needed during the problem-solving tasks and the different perspectives they reveal about student learning (Barrett and Depinet, 1991). Along with case-based reasoning, these approaches are grounded in the belief that real-life reasoning and problem-solving behavior is almost never original and that solutions to new problems are adoptions of previous problem solutions (Kolodner, 1997). Whether this is through the recall of exemplar cases (either representative or contradictory) (Berry and Broadbent, 1988) or by mental model generalizations (Johnson-Laird, 1983) (or scripts [Hudson *et al.*, 1992]) across a number of cases is less clear, as some aspects of strategic reasoning may involve the use of compiled knowledge or implicit memory, i.e., for the most part unconscious (Reder and Schunn, 1996).

The details of the problem-solving approach are often described in terms of the hypothetical-deductive learning cycle where explanations of phenomena are investigated and refined.¹ This cycle is derived from a late stage of intellectual development characterized by students beginning to engage in combinatorial thinking, identification and control of variables, proportional thinking, probabilistic thinking, and correlational thinking. From a cognitive perspective, situations requiring hypothetical-deductive thought often involve a starting condition, a goal condition, and resources to transit between these two cognitive states. In most situations this is an iterative process where intermediate goals (hypotheses) are confirmed/rejected based on the latest information available. If a student were pursuing a particular hypothesis or line of reasoning, the goal of acquiring additional information would be to increase the confidence in the validity of this reasoning chain. Conflicting data, if obtained, would instead

decrease the confidence in the current hypothesis and result in the initiation of a modified search of the problem space.

An important aspect of this model is that students engaged in such activities continually select and revise strategies to optimize the outcomes. Strategies, whether successful or not, are aggregates of multiple cognitive processes including comprehension of the material, search for other relevant information, evaluation of the quality of the information, drawing of appropriate inferences from the information, and use of self-regulation processes that help keep the student on track. Documentation of student strategies at various levels of detail therefore not only can provide evidence of a student's changing understanding of the task, but also can provide experimental evidence of the relative contribution of different cognitive processes to the strategy. Strategies used by students can then become a phenotype, or a proxy so to speak, of the working state of a student's knowledge.

The theoretical design the IMMEX Project uses for investigating students' selection and use of strategies during scientific problem solving is based on extensive work by others (VanLehn, 1996; Schunn *et al.*, 2001; Haider and Frensch, 1996) and can be organized around the following principles.

- Principle 1: Each individual selects the best strategy for them on a particular problem and individuals might vary because of learning in the domain and/or process parameter differences.
- Principle 2: People adapt strategies to changing rates of success.
- Principle 3: Paths of strategy development emerge as students gain experience.
- Principle 4: Improvement in performance is accompanied by an increase in speed and a reduction in the data processed.

THE IMMEX PROBLEM-SOLVING ENVIRONMENT

Given the complexity of strategy selection and use in a problem-solving environment, digital technologies, when appropriately embedded within the context of a discipline and a curriculum, could be useful for documenting student strategies and for using this information to accelerate student achievement. However, the benefits of technology will not be maximized without a theoretical framework(s) around which to design and develop cases. In this article we describe the design, implementation, and validation constructs associated with an on-line problem-solving software system termed IMMEX. For the past 12 years, the IMMEX Project at UCLA has been developing technologies around the broad model of problem solving to probe the development of student understanding in multiple domains (Stevens, 1991; Stevens and Najafi, 1993; Palacio-Cayetano *et al.*, 1999; Underdahl *et al.*, 2001). While originally designed for Intranet delivery, IMMEX cases are now delivered via the Internet (IMMEX, 2003).

The goals of the project have been to

- create a software development environment of the hypothetical-deductive learning cycle,
- create an analytical environment for documenting the formulation and use of strategies, and
- design implementation environments to document student learning with time.

¹For an historical perspective on learning cycles, see Lawson (1995, pp. 155–169).

IMMEX problem solving follows a model of scientific inquiry (Olson and Loucks-Horsley, 2000) and case-based reasoning (Kolodner, 1993) where students are expected to frame a problem from a descriptive scenario, judge what information is relevant, plan a search strategy of the problem space, gather information, and eventually reach a decision that demonstrates understanding. The software was originally designed for medical diagnosis simulations for medical students, a domain that is rich in problem-solving and deductive reasoning (Elstein, 1993). The development of rapid authoring tools along with an intensive program of outreach and professional development, however, has extended the usefulness of IMMEX to elementary, middle, and high schools and to the preservice teacher training curriculum (Palacio-Cayetano *et al.*, 2002).

Students solve IMMEX cases by formulating hypothetical answers, accessing as much information as they feel necessary to test such answers, and then proposing a solution to the problem by selecting an answer from a list of possible answers or by typing in their solution. As students proceed through IMMEX cases, the software records a student's every step as he or she 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 (Stevens, 1991; Stevens *et al.*, 1996; Vendlinski and Stevens, 2002). Currently over 100 problem sets exist in science for middle school to medical school curricula, and over the past 3 years there have been 140,000 student performances.

A Sample IMMEX Case

The sample IMMEX case highlighted in this study is a genetics case involving uncertain parenthood. "True Roots" was designed and created by a team of biology teachers and university faculty to assess student understanding of genetics. In the problem posed, students are introduced to Leucine, who fears that she is a victim of "baby swapping" at birth and begins to conduct a genetic investigation to discover which of five sets of possible parents are truly hers. The students can order tests for blood type, DNA restriction mapping, karyotype, fingerprints, and pedigree charts for both Leucine and each set of parents. For students new to this process, the IMMEX environment also provides students with library resource guides and expert advice enabling students to confirm or reevaluate their interpretation. But because the IMMEX menu structure keeps score to encourage efficiency, use of these items requires decisions that force students to weigh cost versus benefit, decisions that are hotly debated when students work these cases in teams trying to achieve a high score. Using their understanding of inheritance, students include/exclude sets of parents based upon the possibility and impossibility of genetic information being transferred to Leucine and begin to arrive at a solution (Palacio-Cayetano, 1997). There is no predetermined strategy that the students must follow and there are many pathways that students can—and do—take to arrive at the answer. This freedom to navigate information and to form a strategy provides the teacher a great deal of information about student's problem-solving strengths and weaknesses and about their understanding, or lack of, content information (Vendlinski and Stevens, 2001).

Although True Roots was designed for a high school-level understanding of basic genetics principles, it has also been used extensively in the community college and university environments and we have recorded nearly 8,000 performances from these groups. The overall True Roots solution frequency in this dataset is 75.6%.

IMMEX DESIGN FRAMEWORKS

There are two broad design frameworks for the IMMEX problem-solving environment: (1) the design of the problem space and (2) the accumulation, aggregation, and reporting of student performance data. In this regard, our development approach follows the general model of Evidence Centered Design (Mislevy *et al.*, 1999a, 1999b), where the models of the knowledge and skills that we hope students will learn/have learned (student models) are mapped onto behaviors and performance indicators that provide evidence of these constructs (evidence models), which are then designed into tasks that will provide that evidence (task models). We have streamlined this process by involving teachers in all stages of the project including design, development, implementation, and analysis (Palacio-Cayetano *et al.*, 1999; Underdahl *et al.*, 2001).

We are particularly interested in how students develop and use strategies while engaged in problem solving (student model) and we use their sequence of actions while engaged in on-line problem solving (task model) as the evidence from which to infer strategic skill development (evidence model). The problem space itself is constituted by the items and constructs available to the students to navigate with, as well as constraints, such as cost or risks that shape the outcomes that can be expected.

Problem Space Design Framework

There are five specifications that have been used for designing the IMMEX problem spaces. These specifications are crafted in response to the first principle of strategy selection: *Each individual selects the best strategy for them on a particular problem and individuals might vary because of learning in the domain and/or process parameter differences.* This principle suggests that there is significant strategic information in the way a student first perceives and approaches a problem and a problem space, and this may reflect prior knowledge/experience in the domain. To help ensure that meaningful strategic information is captured from the initial and subsequent encounters, we consider (1) the embedded content, (2) the composition of the problem space, (3) the cognitive complexity of the problems, (4) the provision for repeat experiences, and (5) the constraints placed on the problem space.

1. Document the Embedded Content. During the initial design and during the development cycle, the problem spaces and the cases are examined to ensure that the content being presented is appropriate for the intended audience and that the cases are valid and align with relevant instructional and content standards. Teams of teachers, content experts, and educators perform these mappings, often with the input from students, and consider the following issues. First, the content contained in the problem space must be accurate. Next, the scenarios should be real ones where students can become

familiar with the discipline and with possible future job options. This also allows students to practice decision making in a low-stakes environment prior to an internship or laboratory rotation experience. Third, the data must be causally consistent within the cases in that the different items point to the same result of the case. When problem spaces approach or exceed 75 items this can become challenging. Finally, in science problem sets, each menu item on its own often has inherent value, i.e., the molecular weight of a compound, a method for investigating molecular signaling, etc., and it is important that these are contemporary and that background information is provided in the case for students who are less familiar with this information. These mappings are being formalized into on-line searchable XML documents that provide the detailed dimensions of content, standards alignment, and cognitive skills within and across the discipline. The goal of this documentation is to supplement the Dublin Core (2001) and Gateway to Educational Materials metalanguage qualifiers with discipline-specific hierarchies of subject headings to create descriptive XML documents that can satisfy the needs of teachers and researchers alike. An example is available online (IMMEX Problem Set Descriptors, 2003).

2. Construct Challenging and Varied Problem Spaces. This specification addresses the need to accommodate many possible strategic approaches and this is accomplished through the selection of the quantity and quality of items that constitute and structure the problem space (or search space). To avoid linear solutions/thinking and not to limit the diversity of strategies, the problem space should contain many items (30–100). Within this problem space sufficient information should also be present such as library resources, expert opinions, hints, etc., to support students with weaker content

knowledge and to keep them engaged. To promote integrative thinking, no one item should provide an unambiguous solution to the case, and if at all possible, most of the information a student would need to solve the problem should be present in the problem space.

As an example, the True Roots problem space contains 54 data items (available as buttons) that students view in any order they feel appropriate and that contain the information needed to solve the case. Visual representations, or templates, of the problem space can be created in IMMEX where related conceptual items can be grouped together and color-coded (Figure 1).

For this problem set the menu items are arranged by color-code into (1) *reference information*, which helps students to define and interpret concepts and results (upper right); (2) *advice from "experts,"* which is neutral in content but may provide contextual information (middle right); and (3) *research data*, which provide the necessary interpretative data (lower half). These research items consist of different sets of laboratory results from blood typing, restriction fragment length polymorphism (RFLP; commonly known as DNA fingerprinting), and chromosomal karyotype tests, as well as pedigree charts for earlobe attachment inheritance and Punnett squares for analyzing the inheritance of fingerprints.

Efficient student performance and progress in this environment requires that students

- develop an elimination strategy where Leucine is tested for a trait (such as blood typing) and then different parents are eliminated via discordant parent results;
- reduce their use of redundant data, i.e., if parent set 1 has been eliminated by RFLP, they should not retest them; and

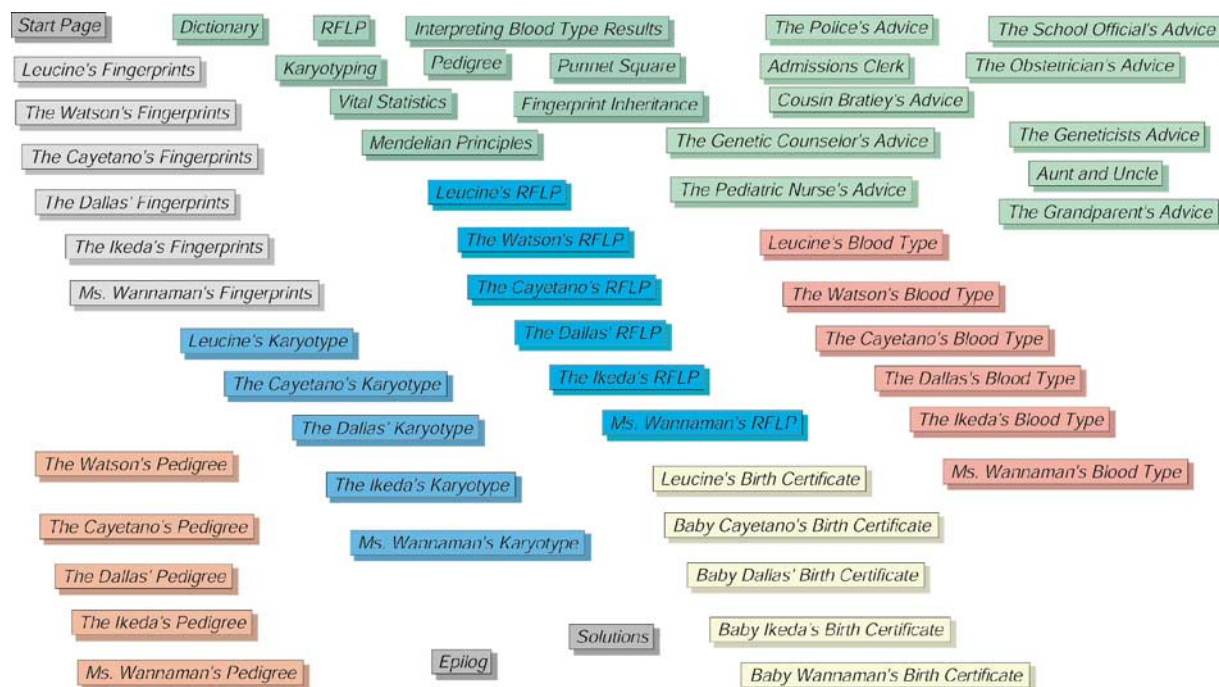


Figure 1. Template representation of the True Roots problem space.

- Optimize their strategies on repeat performances of similar cases by reducing reliance on library material, expert help, etc.

This progression is not intuitive for many students who approach these cases in many different ways and take different trajectories toward becoming efficient problem solvers.

3. Provide and Document Cognitive Complexity. The previous two specifications help developers make a “best guess” about the construction of a particular problem space. Once student performance data are obtained, it is then possible to derive a more refined perspective through research studies.

Given problem spaces of a sufficient size and scope, problem solving within this environment should require an integration of domain knowledge and cognitive process skills, and research studies have shown that IMMEX provides a rich cognitive environment for problem solving. For instance, in a study with 150 undergraduate students, concurrent verbal protocol analysis has indicated that over 90% of student verbalizations while solving a typical IMMEX case could be mapped into 12 distinct cognitive and metacognitive processes including the need for accurate cause–effect inferences, accurate evaluation of information, clarification of gaps in knowledge, monitoring of problem solving behavior, and so forth (Chung *et al.*, 2002). As expected, outside evaluators have also documented that students and teachers perceive the exploratory environment of problem sets like True Roots more as a tool for reasoning and integrating information than as a system for learning new facts (Chen *et al.*, 2001).

4. Provide Repeat Problem-Solving Experiences. The second and third principles of strategy selection and use also suggest that, with time, either within the first simulation, or through multiple performances of parallel tasks, students should be able to learn which strategies are “best” and modify their initial approaches as needed. This implies that opportunities should be available for students to improve and that these changes (or lack thereof) should be documented to show when, and under what conditions, these changes occur.

IMMEX addresses this need by designing problem sets that contain between 5 and 50 different instances (or clones) of a problem that share the same problem space but contain different solutions and different data. These cases can be sequenced by the teachers either randomly or by difficulty, the level being established from prior performances and Item Response Theory analysis. Eventually cases may be automatically staged in response to the current strategic model of the student obtained from prior problem performances. These parallel forms have the further advantages of providing multiple perspectives of a problem space and reducing discussion among students during learning and/or assessment episodes.

The provision of repeat problem-solving experiences can also be extended to a larger design principle that addresses the construction of integrated curricula. The literature suggests the importance of prolonged time/experience with scientific reasoning activities for maximum effect on student performance. To provide such year-round coverage of problem-solving scenarios, we have begun to develop integrated curricular problem sets in chemistry, molecular biology, and biology, each with six to eight different problem sets

(each containing 20+ cases). With these problem-solving curricula students have engaged in over 100 problem-solving experiences over the course of 9 months.

5. Constrain and Extend the Problem Space. For strategic improvement, some measure(s) of success must be a component of the task (Strategic Principle 2), and while such measures can be diverse (a “score,” a “solution,” successfully stabilizing a simulation, submitting an essay to be scored, personal feedback, etc.), some should be included. These components fall into the category of constraints and can have powerful influences on student’s strategies. Constraints are perhaps the least understood of the design specifications. An overall design feature, however, is not to constrain the cases so much, either through hints or by limiting the scope of the question asked, so that the exercise becomes directed and linear.

A common constraint that helps focus student attention on each decision during the performance is to include a cost (or risk or time penalty) for each item requested or for each incorrect “guess” at a solution. A second constraint is the time provided to solve a case and this can be naturally encouraged through a 1- to 2-h course or lab format or can be more structured (i.e., solve three cases in 60 min). Students can also work individually on cases or collaboratively in groups. Evidence from Case *et al.* (2002) suggests that the group environment not only influences the strategies that students use, but also can jog students out of persistent poor strategies. Such constraints can help shape the environment for the students, keeping them engaged and motivated to pursue additional cases (Sedighian, 1997).

Finally it is possible to extend the problems by ending the case in different ways. In most IMMEX cases students choose an answer from a list. On more sociologically based IMMEX problem spaces, the cases end with the students submitting an on-line essay to a question posed in the prologue that is also scored on-line by the faculty (Palacio-Cayetano *et al.*, 2002).

Data Accumulation and Aggregation Specifications

Data collection, aggregation, and reporting should serve the needs of different audiences and link to existing student achievement metrics. IMMEX provides detailed performance data to students and teachers at different levels and in different formats. These reports provide (a) immediate on-line feedback of class and student performance that can link problem-solving efficiency and proficiency to other metrics (AP scores, course grades); (b) visual maps of students’ search of the problem space that reveal problem-solving approaches and strategic changes over time; and (c) research-based predictive models of classroom progress that allow inferences to be made about student progress, which can then suggest points of intervention or allow comparison of problem-solving progress across diverse educational settings.

1. Efficiency and Proficiency Specifications. Student problem-solving proficiency is reflected in the number of cases that students perform and correctly solve, while efficiency takes into account the time spent on the case(s) and the resulting scores. On-line, teachers can view graphs of these metrics that plot the number of problems completed versus the number solved and then drill down to look at this

ratio for individual students. These further hyperlink to individual performances where student progress throughout the course is documented. Examples of these reporting formats can be found at IMMEX (<http://www.immex.ucla.edu>).

These reporting features are therefore useful for providing a quick snapshot of student performance. Often, however, teachers wish to dig deeper into the foundations of their students thinking and the data reporting features of the strategic specifications equip teachers with the tools to do so.

2. Strategic Specifications. A schema for the collection, aggregation, and reporting of strategic performance data is shown in Figure 2. The data collection begins with the recording of each item selected (Figure 2, I) by students as they navigate the problem space in response to the problem posed. While humans perform many parallel mental operations, our explicit actions are mostly serial and deliberate. The microinformation of students' individual choices during problem solving would therefore be expected to contain strategic information, and in fact, when combined with time latencies, this information also provides some information about the distribution of effort during the problem-solving episode. For instance, Paek (2002) demonstrated that on a series of IMMEX mathematics problems, students solving these cases spent 35% more time

on the first step of the problem than students who missed the problems. This is analogous to the finding that experts spend proportionally more time in framing a problem than do novices (Baxter and Glaser, 1997).

While a single student action during the performance is occasionally informative in IMMEX problem solving (such as "guessing" when a student chooses to solve a case as an initial move and without viewing any information), experience suggests that a sequence of actions is more consistently revealing. These sequences are seldom random, and while students may eventually look at all the information contained in a problem space, they will often view menu items sequentially along pathways that share some common features rather than following entirely haphazard approaches. We believe that the paths that students employ while completing an IMMEX case provide evidence of a (complex) 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.

This sequence and search information is captured in the completed performance (Figure 2, II), which is the next level of data aggregation. Here, students have searched the problem space, made a decision(s) about the embedded problem, and received feedback as to whether or not they were

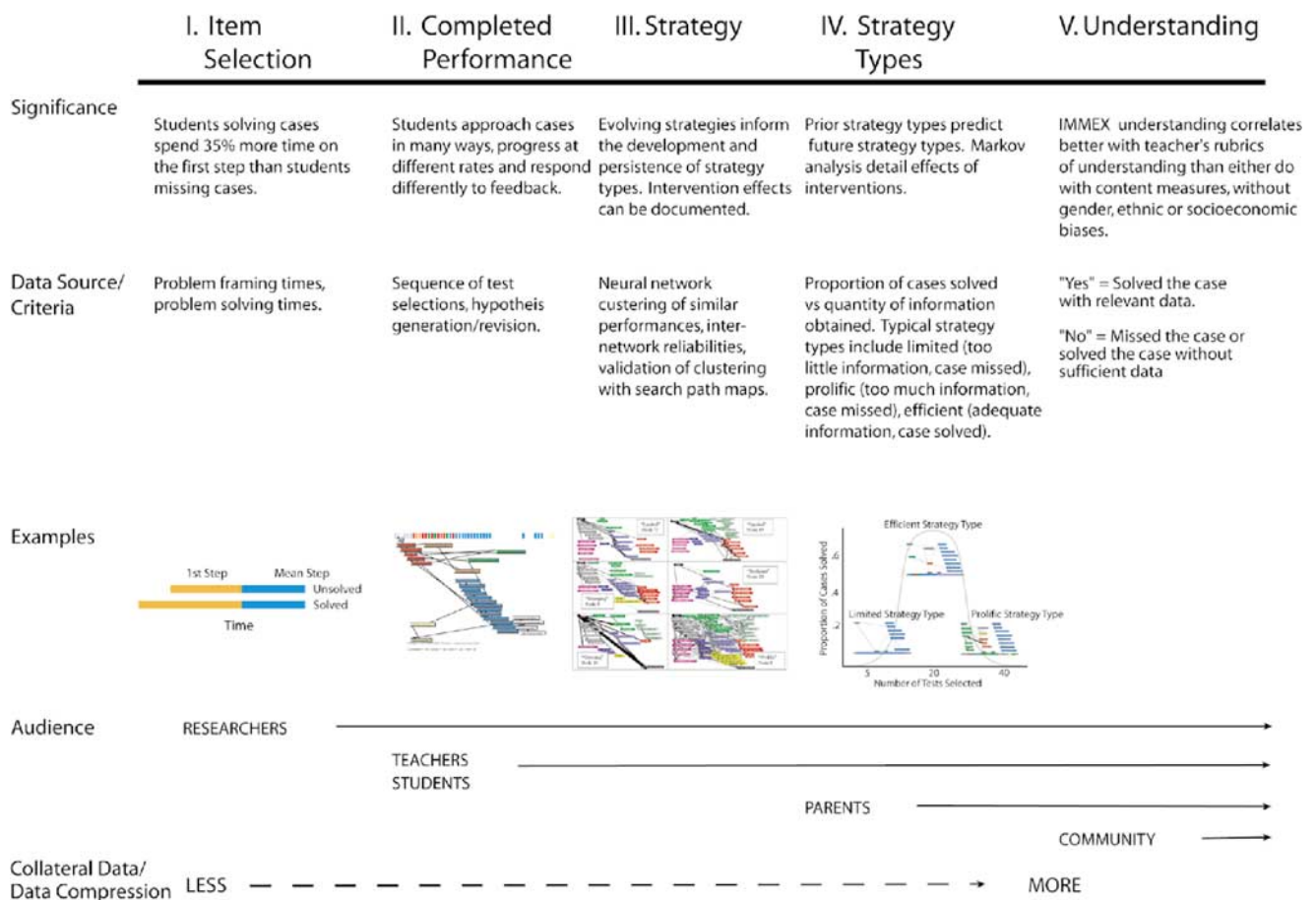


Figure 2. IMMEX performance reporting constructs.

successful. To help make these complex data understandable to teachers and students, we generate visual maps of their search through each problem space. This search path mapping software (Stevens, 1991) displays each data item selected by the student (represented by the color-coded rectangles arranged by content, concepts, or type), the sequence in which the information was selected (indicated by the line traveling from the left corner to the center of the rectangle), and the time spent on each concept (see Figure 3).

Because these maps are information-rich they can be used in multiple ways. For researchers and teachers who are involved in authoring the case, these maps provide a validity check to ensure that the case is performing as intended. These maps can also be provided to students to encourage reflection and support their own analysis of their performance and progress (Lawton, 1998). Classroom interventions are further suggested by comparing the strategies used by the students who solved or missed the case. Finally, search path maps can be important for examining the more metacognitive aspects of problem solving such as persistence, elimination of alternative hypotheses, efficiency, confidence, and certainty. Profes-

sional development activities for faculty and teachers center around the use of search path mapping techniques for these types of educational interventions.

DYNAMICS OF STRATEGY CHOICE AND CHANGE

Close examination of many of these maps has revealed that the use of different strategies (Figure 2, III) by students when solving a series of True Roots cases is a dynamic process and that there is extensive restructuring of the strategic approach as experience is developed. While this was originally identified by examination of hundreds of search path maps, over the past years we have used artificial neural technologies to begin to automate this process. Details of these techniques have been reported previously (Stevens and Najafi, 1991; Hurst *et al.*, 1999; Vendlinski and Stevens, 2002) and are outside the scope of this article.

For the True Roots problem set, we have studied these dynamics in a series of controlled experiments involving community college and university students. Using over 500

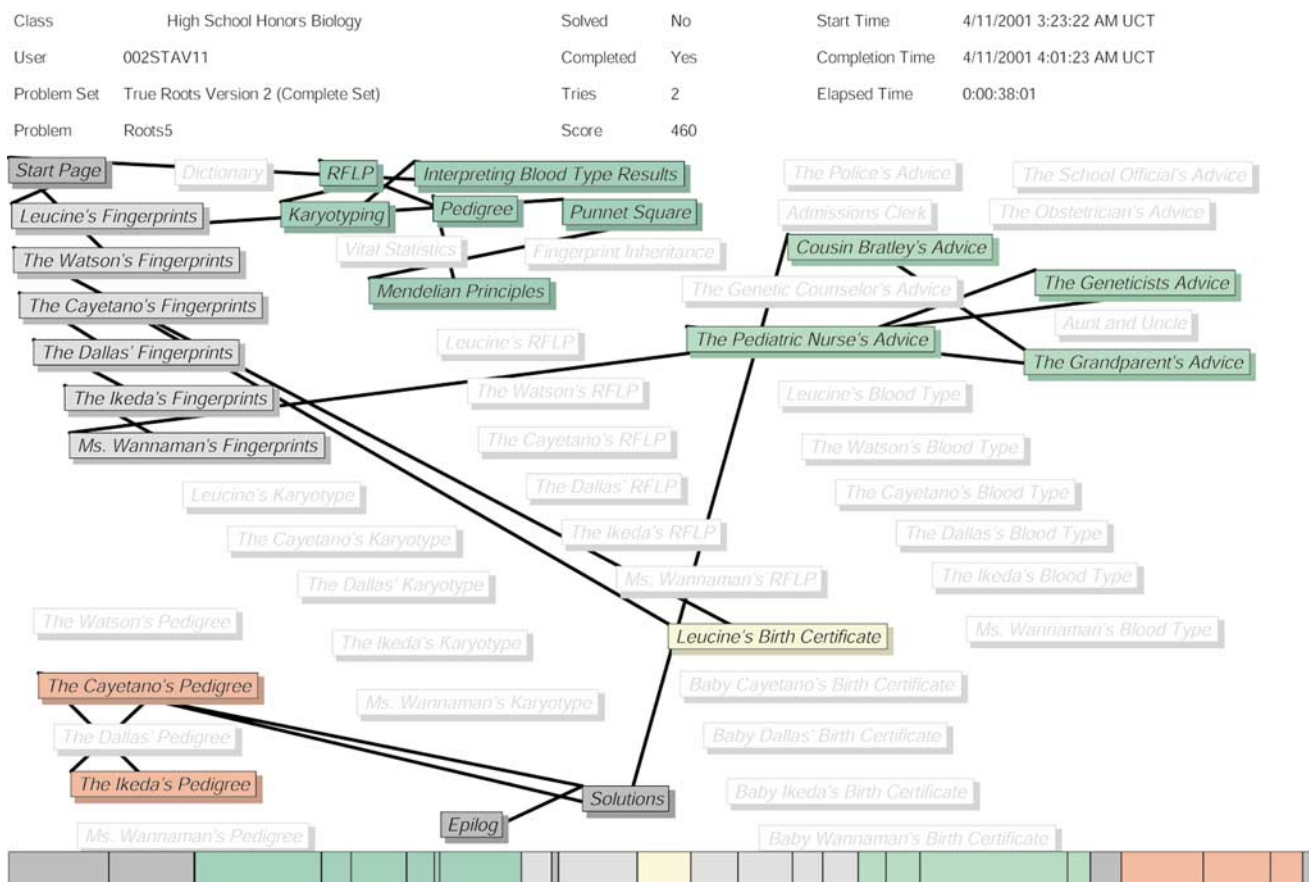


Figure 3. Sample search path map. The performance of student 002STAV11 has been overlaid on the True Roots template with lines connecting the sequence of test selection. The lines go from the upper left-hand corner of the “from” test to the lower center of the “to” test. Items not selected are shown in transparent gray. At the top are the overall statistics of the case performance. At the bottom is a time line where the colors of the squares link to the colors of the items selected and the width of each rectangle is proportional to the time spent on that item. These maps are immediately available to teachers and students at the completion of a problem.

Table 1. “True Roots” strategy type descriptions

Strategy type	Description
Prolific	Categorized by a thorough search of the problem space including the five relevant concept/data domains as well as resource and conjecture menu items
Redundant	Shows a reduction in the use of resource and conjecture items but maintains a comprehensive search of the data domains. Reordering tests for parents who have already been eliminated by prior testing is also characteristic of this type of search
Efficient	Depicts searches that access only that information needed to solve the case, and do not order tests for parents who have been eliminated through prior test selections
Limited/guessing	Demonstrates premature closure of the problem as indicated by not gathering enough information to solve the case conclusively

performances from these students, we have aggregated the most common strategies into four main categories or *strategy types* (Figure 2, IV), which appear to be the most representative of the cognitive requirements of the task. These are defined by the criteria in Table 1. The dynamic changes in the use of these strategies as students perform the five cases in the True Roots problem sets are detailed below.

Prolific Strategies

Typically, students who are being introduced into a new IMMEX problem-solving environment begin with a broad investigation of relevant and irrelevant information coupled with the recognition and/or learning of key concepts. It is not unusual to see students use this strategy on the first case of a problem set as they are exploring and defining the problem space, collectively called framing. However, if learning is occurring, then we would expect that with practice on multiple cases, task-relevant items will be separated from task-redundant information and the most relevant aspects of the task will be focused on (Haider and Frensch, 1996). This should result in the abandonment of the prolific strategy, an increase in the solve rate, and a decrease in the amount of time spent per case. As shown in Figure 4 most students began with a prolific strategy type, and on subsequent cases the use of this strategy type declined. By the fourth case, the students still using this strategy were particularly at risk for missing the solution.

Redundant Strategy Type

When students can distinguish relevant and irrelevant information, they begin selectively visiting the content domains they are most comfortable with to eliminate possible parents. In this regard, it is not unusual for students to start to develop favorite sets of test items, such as blood typing or fingerprinting, that they will consistently embed into their

strategies. The selection of a favorite repertoire of tests is complex and, from the verbal protocols, is at least in part influenced by familiarity with the terms from prior experiences (TV, newspapers, science courses) and/or by the format in which the data are displayed. For instance, pattern matching between RFLP digests seems easier for some students than does antibody agglutination, which has more complex fundamentals.

The redundant strategy type appears to be an intermediate stage in the development of efficient strategies in that the student understands the need for comparing the child Leucine with the different parents across a particular test to eliminate possible parents. The strategy is not optimized, however, as the same parents are often retested multiple times with different laboratory tests even though they should have been eliminated.

As shown in Figure 5, the redundant strategy was the second most frequent starting strategy and students using it initially had a relatively poor success rate. The frequency of use of this strategy increased on the second case (with a corresponding decrease in the prolific strategy) and then progressively declined.

Efficient Strategy Type

The solve rate for the efficient strategy type is high (80%) and consistent (Figure 6). Few students in this test set understood the problem space well enough on the first case to employ an efficient strategy. The use of this strategy slowly increased with each case attempt, peaking at case 4.

Limited Strategy Type

Students who are attempting to solve the case with insufficient information (at least on the early problem sets) are grouped into the limited strategy type (Figure 7). The frequency of use of this strategy is highly implementation-site-dependent, with some high school classes showing a preponderance of guessing strategies. In this tightly controlled experimental situation, however, few of these more advanced students initially used a limited strategy. Paradoxically, the use of this strategy increased on subsequent cases and was accompanied by a high solution frequency. Retrospective analysis of their verbal protocols indicated that some of the students were going outside the scope of the problem space and learning a higher strategy (i.e., knowing there are only five parents and eliminating those they had experienced in prior cases). Following this result, we suggest that a problem set contain a minimum of 10+ cases to restrict the use of this particular strategy. The decrease in all the numbers on the fifth case is due to students not completing the five problem sets in the allotted time.

An example of such strategic transitions for a single high-school student is shown in a search path map format in Figure 8. Here a student on the first True Roots case performed a thorough exploration of the problem, spent 65 min on it, and did not solve it correctly. Several days later, the student spent 29 min on the second case and, at this time, had discarded the use of library resources. While laboratory data were being accessed, some of them were redundant. By the third case the student's strategy had become efficient, requiring only 6.5 min to solve the case.

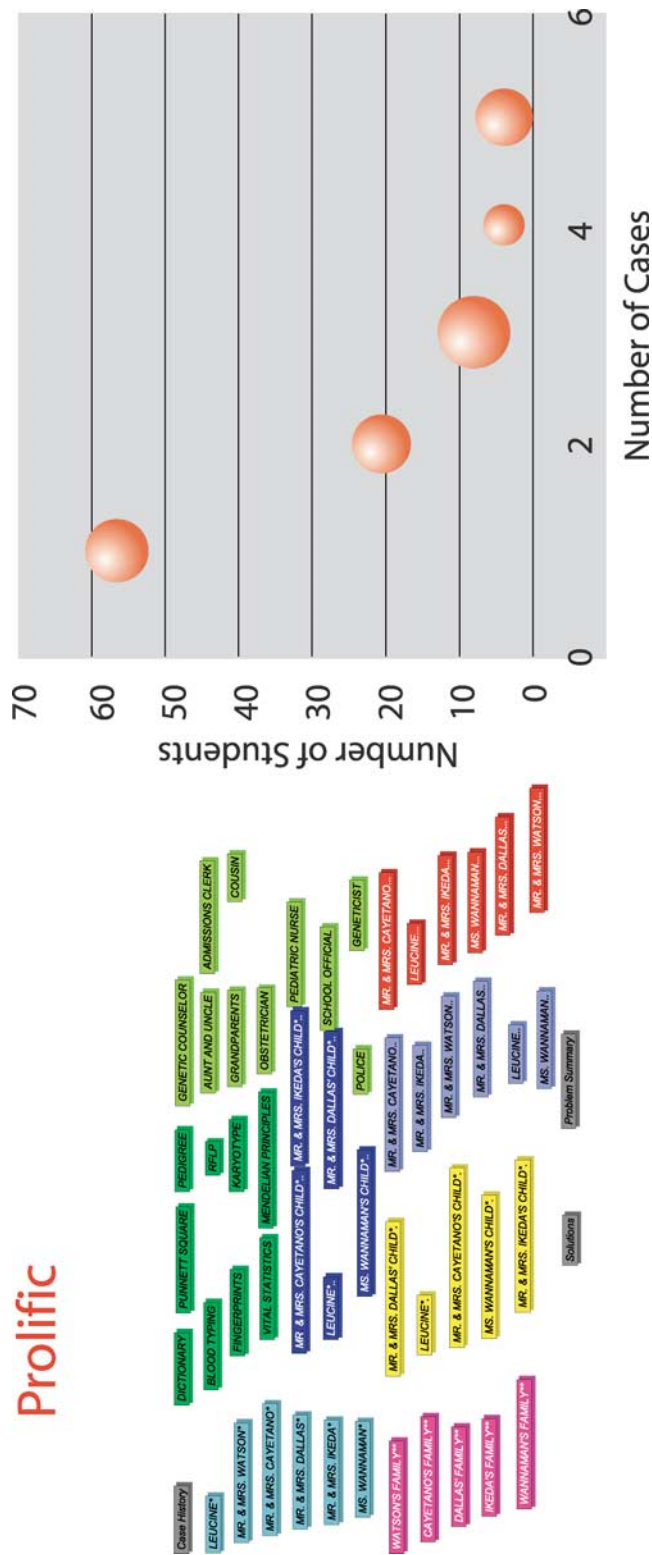


Figure 4. Dynamics of the the prolific strategy type. The number of students using the prolific strategy type (Y-axis) is plotted vs. the case performed. The size of the circles is proportional to the solution frequency.



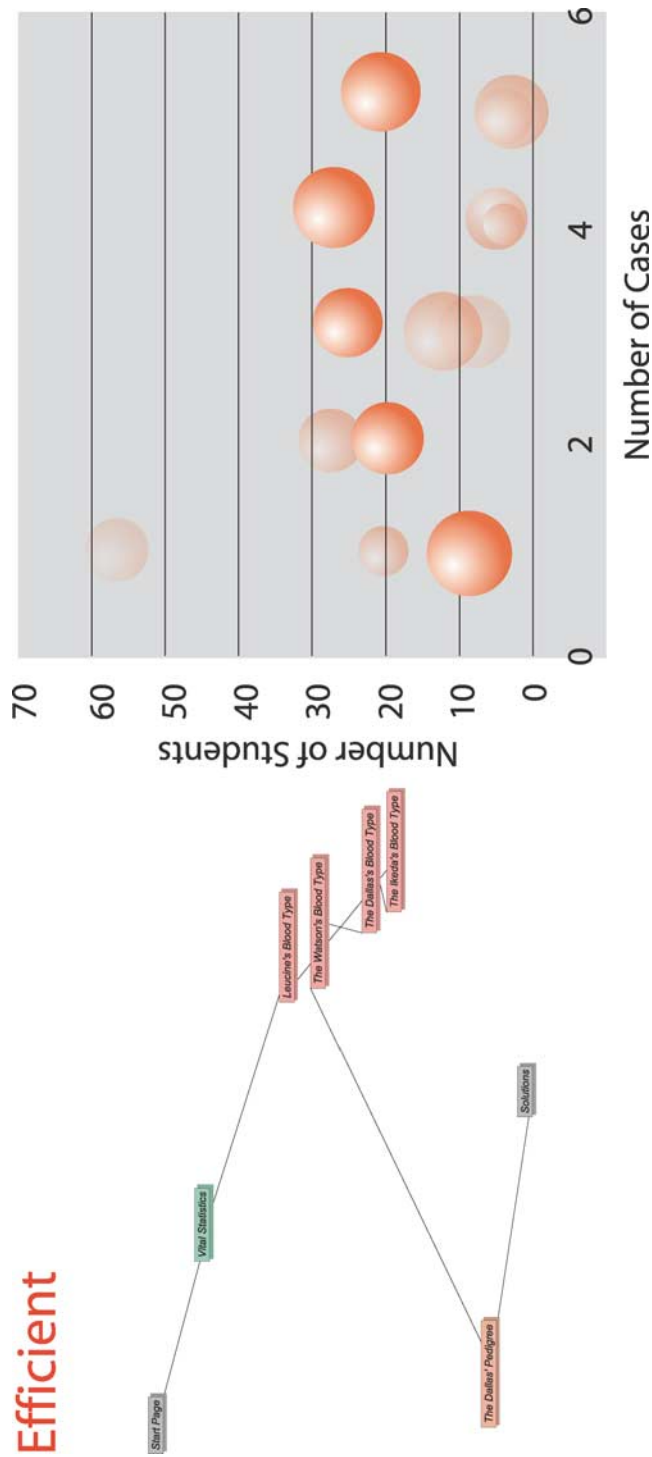


Figure 6. Dynamics of the efficient strategy type. The number of students using the efficient strategy type (Y-axis) is plotted vs. the case performed. The size of the circles is proportional to the solution frequency.

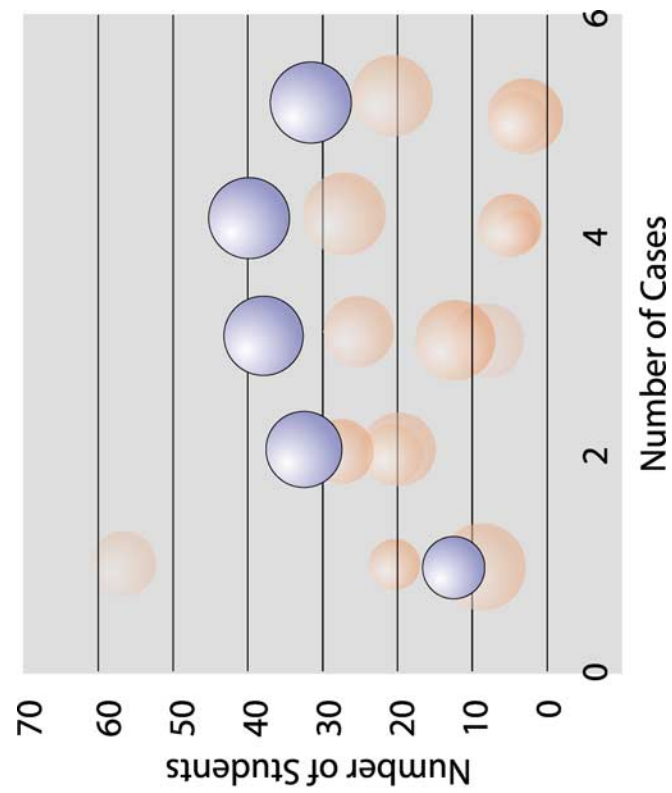


Figure 7. Dynamics of the limited strategy type. The number of students using the limited strategy type (Y-axis) is plotted vs. the case performed. The size of the circles is proportional to the solution frequency.



Second Performance (29 minutes)



Figure 8. Sequence of performances of one student on three True Roots cases.

VALIDATING THE CONSTRUCTS OF STRATEGIES AND STRATEGY TYPES

Previous validation studies (Kanowith-Klein *et al.*, 1998; Palacio-Cayetano, 1998; Chung *et al.*, 2002) have shown that IMMEX in general and True Roots in particular are

- cognitively complex,
- engaging and motivating for students and teachers, and
- capable of being implemented in a wide number of disciplines and settings.

We have also shown that it is possible to create cases with the same level of difficulty (as measured by Item Response Theory) as well as with predictable levels of difficulty (Vendlinski and Stevens, 2002).

The following studies were conducted to examine whether the True Roots problem set behaved in a predictable way when developed and analyzed according to the above framework. Our hypotheses were (1) that students of different abilities would approach and solve these cases with strategies consistent with these abilities, (2) that students with higher measures of scientific reasoning would employ more effective strategies, and (3) that students who are developing effective strategies would show greater improvement in speed with practice. These studies address the concurrent validity (the ability to distinguish between groups it should theoretically be able to distinguish between) and convergent validity (the degree to which the construct converged on other constructs).

Relating Ability to Problem-Solving Strategic Approaches

The student populations that could potentially provide us with varying degrees of ability were recruited from a university freshman class ($n = 94$) and entering students from a nearby community college ($n = 105$). We predicted that by tapping two different levels of the institutionalized selection process, we would gain the maximum variability within the 17- to 18-year-old range to link to the strategic trends that emerge along this spectrum.

The ability of the two groups of students was also significantly different on the task itself: the university students were above (82%) the average solution frequency of the 8,000+ performances of True Roots in our database (76%), while the community-college students were below this average, at 67% ($p < .001$). Differences were also seen in the way that the students approached the cases. The community-college students accessed significantly more resource ($p < .001$) and conjecture ($p < .001$) items than did the university-level students.

A more detailed indicator of ability level differences between the community-college level and the university participants involved the linkage between what students *said* during the concurrent verbalization and what they *did* simultaneously on the IMMEX task. The community-college students used more “simple statements”—repeating of text from the item ordered indicative of nonanalysis—($p < .001$), had fewer “correct cause and effect” statements ($p = .014$), were less able to evaluate information correctly ($p = .003$), and asked fewer questions during decision making, which indicates lack of hypothesis building ($p = .026$). Moreover, the university students were more likely to make statements that clarified gaps in knowledge ($p = .012$) and ar-

Table 2. Strategy type use by community-college and university students: Strategy type by ability level

Strategy type	Ability level		Total
	University	Community college	
Prolific			
Count	58	98	156
Expected count	73.8	82.2	156.0
Redundant			
Count	91	105	196
Expected count	92.7	103.3	196.0
Efficient			
Count	218	181	399
Expected count	188.7	210.3	399.0
Limited			
Count	34	63	97
Expected count	45.9	51.1	97.0
Total			
Count	401	447	848
Expected count	401.0	447.0	848.0

ticulate judgments of information relevancy ($p < .001$), while the community-college students verbalized more awareness of task goals and their progress—or lack thereof—toward achieving those goals ($p < .001$).

While both university and community-college students used all four strategy types defined in Table 1 across the five cases, the university students used a higher proportion of Efficient strategies, while the community college students used more Prolific and Limited strategy types (Table 2) (Pearson $\chi^2 = 44.2$, $p < .001$).

The dynamic changes in the use of prolific and efficient strategies, which showed the greatest disparity between the community-college and the university students, were then examined across the five cases. As shown in Figure 9, the university students appeared to be a step ahead of the community-college students in the abandonment of prolific

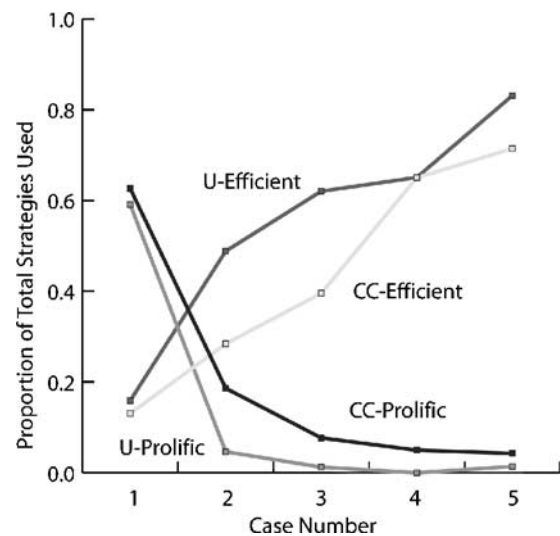


Figure 9. Dynamics of prolific and efficient strategy type use by community-college (CC) and university (U) students. The number of students using the prolific and efficient strategies is plotted vs. the case performed.

Table 3. Strategy type use by students with different CTSR scores: University and community-college strategy types by CTSR scores

Strategy type	100–80	79–60	59–50	49–30	29–0	Total
Prolific						
Count	31	46	31	34	26	168
Expected count	40.1	51.2	29.2	29.4	17.9	168.0
Redundant						
Count	24	48	20	28	21	141
Expected count	33.7	43.0	24.5	24.7	15.1	141.0
Efficient						
Count	129	140	78	61	26	434
Expected count	103.7	132.3	75.5	76.1	46.4	434.0
Limited						
Count	15	20	16	23	16	90
Expected count	21.5	27.4	15.7	15.8	9.6	90.0
Total						
Count	199	254	145	146	89	833
Expected count	199.0	254.0	145.0	146.0	89.0	833.0

and redundant strategies and between one and two steps in the adoption of efficient strategies.

To provide additional evidence of the difference in student ability, both student populations completed the Classroom Test of Scientific Reasoning (CTSR) (Lawson, 2000) as an independent measure of scientific reasoning ability level. The modified version of the CTSR (22 multiple-choice items) is designed to assess a student's ability to separate variables, conserve weight and volume, and use proportional logic as well as combinational reasoning and correlations. The university students had a mean score of 71%, while the community-college students had a mean score of 49%, and this disparity was significant ($t = 7.844$, $p < .001$).

Next, the students' use of different strategy types was measured as a function of their CTSR scores. As shown below, those students who scored on the lower end of the CTSR used more Prolific, Redundant, and Limited strategies, while those students who scored on the upper end of the CTSR used more Efficient strategies than expected (Table 3).

Strategy Improvement Reduces Performance Time

A final hypothesis was that as students begin to develop successful strategies the time needed to solve each subsequent case should decrease. This hypothesis derives from the power law of practice as it relates to the use of strategies on complex tasks (Delaney *et al.*, 1998). Practice on a task almost always improves performance, both by reducing the number of errors and by reducing the time to perform the task. With True Roots, the reduction of errors is reflected in the refinement of the strategies toward more efficient ones. This should also be reflected in a decrease in the time needed to solve the case.

The performance times of a group of 15 community-college students that solved all of their True Roots cases and showed strategic improvement and a group of 15 students that missed at least three of the five cases are shown in Figure 10. The high-solve group spent more time on the first case than the low-solve group and then showed a more rapid ($p < .01$) decrease in the solution time on subsequent cases.

Combined, these data also suggest that the search path mapping and artificial neural network analytic tools embedded in the IMMEX system can discern differences in the level of expertise of the user. This, in fact, has been confirmed in the domain of medical immunology, where performances of expert clinicians could be distinguished from those of medical students (Stevens *et al.*, 1996).

DISCUSSION

Linking the IMMEX Design Environment with Other Multimedia Frameworks

The previous sections have described design frameworks for investigating student strategies and have provided validation evidence. Similar analyses have been/are being conducted for high-school and university chemistry problem sets (Vendlinski and Stevens, 2002) and middle-school science and medical education (Casillas *et al.*, 1999) applications. We therefore believe that the principles can be generalized to many, if not most, of the 100+ existing IMMEX problem sets.

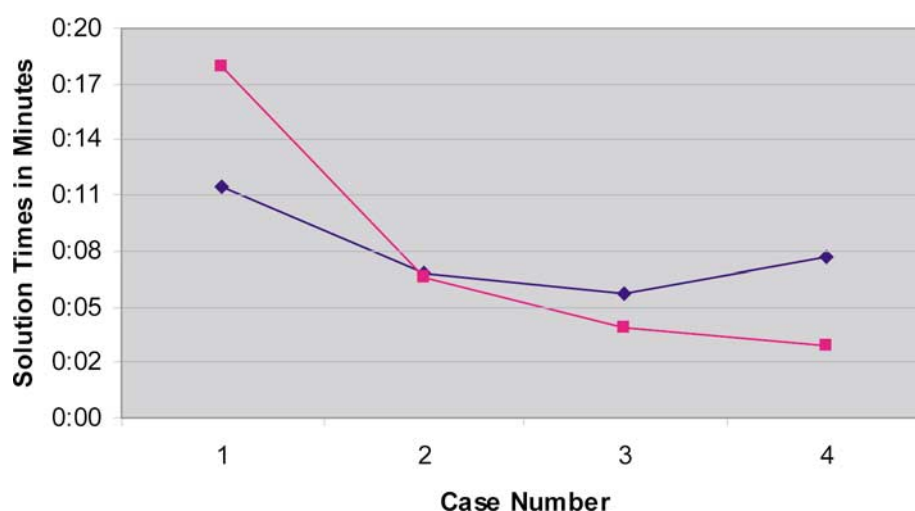


Figure 10. Problem solution times as a function of practice. The mean solution time for a group of 15 students who solved at least four True Roots cases (red) and a group that missed at least four cases (blue) is plotted vs. the case number.

In this final section we situate the IMMEX design framework within the context of other frameworks that can, and do, influence the strategies that students use while engaged in on-line learning or assessment. Such frameworks include the instructional presentation modes and issues of implementation (classroom vs. homework assignments, individual vs. group performance, and teaching and learning vs. assessment).

Instructional Presentation Modes. One complementary topic is the theory of multimedia learning that addresses different instructional presentation modes. This theory, elaborated by Mayer and associates (Mayer, 2000; Mayer *et al.*, 2001) and others (Swiller and Chandler, 1994), is based around six different principles that relate to the most effective approaches for presenting material through verbal and nonverbal modes. These results suggest that students learn better (1) when verbal information is presented verbally as opposed to on-screen text (Modality Principle), (2) when they do not need to split their attention between multiple sources of information (Split Attention Principle), (3) when visual and spoken materials are temporally synchronized (Temporal Contiguity Principle), (4) when on-screen text and visual materials are physically integrated (Spatial Contiguity Principle), (5) when students are simultaneously presented animation and narration then animation, narration, and text (Redundancy Principle), and (6) when extraneous material is excluded rather than included in the multimedia explanations (Coherence Principle).

Currently most IMMEX simulations present data in a more static and concise format than those of Mayer *et al.* and we have not explored the influence of these different factors on student strategic approaches, although we would expect them to be significant. One approach we are exploring is to begin to replace static data items with more dynamic applets where students need to manipulate variables in a dynamic model in order to derive a piece of information relating to the problem. Nevertheless, irrespective of the presentation format(s) of the investigative data, for the investigation of student strategies each item should require/help each student (1) to understand the data being presented and (2) to relate these new data to information previously acquired and to the current hypothesis.

The above theory of Mayer *et al.* does raise the questions of how much practice is required to “master” the IMMEX interface and how much of student improvement is based on simply learning to navigate this interface. The interface used in IMMEX, where there are cascading menus of potential data items, has undergone steady refinement since the early 1990s, when the program was Microsoft Windows-based. The consistency of the problem presentation has had the advantage that when students engage in year-round problem solving, with multiple problem sets (i.e., some in chemistry, some in biology, etc.), they do not need to learn a new interface each time, and they can more effectively focus their efforts on the problem at hand. But what about students’ first use of IMMEX? Currently students beginning at Grade 6 are performing the cases and they navigate the menu structure with little difficulty, and repeated student surveys across a range of grades and disciplines seldom mention user interface issues. More concretely, during two separate verbal protocol studies with nearly 100 students each, students were given the opportunity to practice with the interface on a middle-school

science problem until they felt comfortable before performing a series of True Roots cases. Most students in both of these groups felt that they were comfortable with the interface after the performance of a single practice case. Furthermore, within the verbal protocols, comments regarding the user interface were rare (<1% of utterances) (Chung *et al.*, 2003).

Implementation Issues. There are also the implementation issues of infrastructure and teacher preparation that can influence the distribution of strategies that students use on the IMMEX simulations. Both of these are complex issues. Regarding infrastructure, data for validating this study were obtained from controlled conditions conducted by researchers and, while restricted to a 1-h time frame, nevertheless was administered in a research setting with a high-speed Internet connection, conditions not often found in most classrooms or in curricula where student contact time is limited. In response to these constraints, users of IMMEX are exploring different implementation modalities. For instance, some faculty and teachers have assigned the cases as homework rather than use classroom time. This has provided students more problem-solving opportunities (up to 100 over the course of a year) and the on-line data reporting tools have let teachers follow the progress of different classes and students on-line (Stevens *et al.*, 1999). Other teachers have begun to explore the challenges of using these cases for student assessment or have explored the effects of collaborative learning.

We have examined the performances of hundreds of the 8,000+ field-based student strategies of True Roots using the same rubrics and artificial neural network clusters developed for this study and have obtained evidence regarding individual student strategic performance and progress consistent with the research-based results presented here. What is more noteworthy, however, occurs at the aggregate classroom level, where there is significant variance of strategy usage across different schools and classrooms. At one extreme there are classrooms that show a distribution of strategy use similar to that reported here. At the other extreme there are classrooms where the majority of the students use an unproductive limited approach. These results suggest implementation effects reflecting infrastructure and/or teacher preparation differences (Palacio-Cayetano *et al.*, in preparation).

Finally, while technology infrastructures continue to improve (and be adequate for the more simple IMMEX presentation formats), the preparation required by both faculty and students for the implementation of these technologies and the analysis of student performance is not yet formalized. This is a broad topic, impossible to discuss more than summarily in this article, and in fact, an entire program of the U.S. Department of Education has been devoted to training the educational community on these issues (PT³; Preparing Tomorrow’s Teachers to Use Technology). This has resulted in a comprehensive series of standards and guiding implementation frameworks, many of which can be found on-line (International Society for Technology in Education, 2003).

Accessing Materials

IMMEX problem sets can be accessed by teachers and educators from the IMMEX Project’s home page (<http://www.immex.ucla.edu>). To begin, users should select the link

"New User Registration," which will generate a unique ID and provide access to dozens of sample cases including True Roots. Once cases are identified for classroom use, faculty can request to stage a class by selecting the link "Staging Request Form." Here, the faculty can enter the name of the course, the number of students, dates the cases should be available, etc.

Detailed instructions for accessing and using the data reporting tools can be found at <http://www.immex.ucla.edu/IMMEXMainFrame.htm>. Additional manuscripts and background information can be found at "Publications" at <http://www.immex.ucla.edu/ProjMainFrame.htm>. There is no charge for using the IMMEX cases.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Science Foundation (NSF-ROLE 0231995, DUE Award 0126050, ESE 9453918), the PT³ Program of the U.S. Department of Education (Implementation Grant P342A-990532), and the Howard Hughes Medical Institute Precollege Initiative. The author thanks Dr. Terry Vendlinski as well as Jennifer Underdahl of IMMEX, Dr. Greg Chung and Linda DeVries of CRESST, Dr. Melanie Cooper and Edward Case of Clemson, and Dr. Marcia Sprang of Esperanza High School for their outstanding collaborative efforts.

REFERENCES

Anderson, J.D. (1980). *Cognitive Psychology and Its Implications*, San Francisco: W.H. Freeman.

Barrett, G.V., and Depinet, R.L. (1991). A reconsideration of testing for competence rather than for intelligence. *Am. Psychol.* 46, 1012–1024.

Baxter, G.P., and Glaser, R. (1997). *An Approach to Analyzing the Cognitive Complexity of Science Performance Assessments*, Los Angeles: National Center for Research on Evaluation, Standards and Student Testing (CRESST), Center for the Study of Evaluation, University of California, Los Angeles.

Berry, D.C., and Broadbent, D.E. (1988). Interactive tasks and the implicit-explicit distinction. *Br. J. Psychol.* 79, 251–272.

Case, E.L., Cooper, M., Stevens, R., and Vendlinski, T. (2002). Using the IMMEX system to teach and assess problem solving in general chemistry. ACS Annual Meeting Presentation, Orlando, FL. http://www.acs.org/portal/PersonalScheduler/EventView.jsp?paper_key=200723&session_key=34422.

Casillas, A.M., Clyman, S.G., Fan, Y.V., and Stevens, R.H. (1999). Exploring alternative models of complex patient management with artificial neural networks. *Adv. Health Sci. Educ.* 1, 1–19.

Chen, E., et al. (2001). How teachers use IMMEX in the classroom. <http://www.immex.ucla.edu/TopMenu/WhatsNew/EvaluationForTeachers.PDF>.

Chung, G.R., de Vries, L.F., Cheak, A.M., Stevens, R.H., and Bewley, W.L. (2002). Cognitive process validation of an online problem solving assessment. *Comput. Hum. Behav.* 18, 669–684.

Delaney, P.F., Reder, L.M., Staszewski, J.J., and Ritter, F.E. (1998). The Strategy-specific nature of improvement: The power law applies by strategy within task. *Psychol. Sci.* 9, 1–7.

Dublin Core Qualifiers (2001). [http://purl.org/dc/Gateway to Educational Materials](http://purl.org/dc/Gateway%20to%20Educational%20Materials) (2001). <http://www.geminfo.org/Workbench>.

Elstein, A.S. (1993). Beyond multiple-choice questions and essays: The need for a new way to assess clinical competence. *Acad. Med.* 68, 244–249.

Haider, H., and Frensch, P. (1996). The role of information reduction in skill acquisition. *Cognit. Psychol.* 30, 304–337.

Hudson, J., Fivush, R., and Kuebli, J. (1992). Scripts and episodes: The development of event memory. *Appl. Cognit. Psychol.* 6, 625–636.

IMMEX Problem Set Descriptors (2003). <http://www.immex.ucla.edu/SubjectHeadingDocs/hazmat.htm>.

IMMEX Project Web Site (2003). <http://www.immex.ucla.edu>.

International Society for Technology in Education (2003). Center for Applied Research in Educational Technology (CARET). <http://caret.iste.org>.

Johnson, M.A., and Lawson, A.E. (1999). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes. *J. Res. Sci. Teach.* 35, 89–103.

Johnson-Laird, P.N. (1983). *Mental Models*, Cambridge, UK: Cambridge University Press.

Kanowitz-Klein, Burch, C., and Stevens, R. (1998). Sleuthing for science. *Natl. Staff Dev. Council J. Staff Dev.* 19(3), 48–53.

Kolodner, J.L. (1993). *Case-Based Reasoning*. San Mateo, CA: Morgan Kaufman.

Kolodner, J.L. (1997). Educational implications of analogy. A view from case-based reasoning. *Am. Psychol.* 52, 57–66.

Krijik, J., and Haney, R. (1987). Proportional reasoning and achievement in high school chemistry. *School Sci. Math.* 70, 813–820.

Lawson, A.E. (1995). *Science Teaching and the Development of Thinking*. Belmont, CA: Wadsworth.

Lawson, A.E. (2000). Classroom test of scientific reasoning [multiple choice version]. Based on the development and validation of the classroom test of formal reasoning. *J. Res. Sci. Teach.* 15, 11–24.

Lawson, A.E., Clark, B., Cramer-Meldrum, E., Falconer, K.A., Sequist, J.M., and Kwon, Y.-J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? *J. Res. Sci. Teach.* 37, 81–101.

Lawton, M. (1998). Making the most of assessments. *Educational Week on the Web*, October. <http://www.edweek.org/sreports/tc98/cs/cs9.htm>.

Mayer, R.E., and Moreno, R. (2000). A learner-centered approach to multimedia explanations: Deriving instructional design principles from cognitive theory. *Interactive multimedia. Electronic Journal of Computer-Enhanced Learning*. <http://imej.wfu.edu/articles/2000/2/05/index.asp>.

Mayer, R.E., Heiser, J., and Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *J. Educ. Psychol.* 93, 187–198.

Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Almond, R.G., and Johnson, L. (1999a). A cognitive task analysis, with implications for designing a simulation-based assessment system. *Comput. Hum. Behav.* 15, 335–374.

Mislevy, R.J., Almond, R.G., Yan, D., and Steinberg, L.S. (1999b). Bayes nets in educational assessment: Where do the numbers come from? In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, eds. K.B. Laskey and H. Prade, San Francisco: Morgan Kaufmann.

Novak, J., and Gowin, D. (1984). *Learning How to Learn*, New York: Cambridge University Press.

Olson, A., and Loucks-Horsley, S., eds. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*, Washington, DC: National Academy Press.

Paek, P. (2002). Problem solving strategies and metacognitive skills on SAT mathematics items. Unpublished thesis, University of California, Berkeley.

Palacio-Cayetano, J. (1998). Unpublished thesis, University of Southern California, Los Angeles.

- Palacio-Cayetano, J., Kanowith-Klein, S., and Stevens, R. (1999). UCLA's outreach program of science education in the Los Angeles schools. *Acad. Med.* 7(4), 348–351.
- Palacio-Cayetano, J., Schmier, S., and Dexter, S. (2002). Experience counts: Comparing preservice and inservice students technology integration decisions. Society for Information Technology and Teacher Education International Conference (SITE 2002). Nashville, TN, March 18–23.
- Pelligrino, J., Chudowski, N., and Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academies Press.
- Reder, L., and Schuun, C. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit memory and learning. In: *Implicit Memory and Cognition*, Mahwah, NJ: Lawrence Erlbaum Associates. Mahwah, NJ.
- Schuun, C.D., and Anderson, J.R. (2002). The generality/specificity of expertise in scientific reasoning. *Cognit. Sci.* 23(3), 337–370.
- Schuun, C.D., Lovett, M.C., and Reder, L.M. (2001). Awareness and working memory in strategy adaptivity. *Memory Cognit.* 29(2), 254–266.
- Sedighian, K. (1997). Challenge-driven learning: A model for children's multimedia mathematics learning environments. ED-MEDIA 97: World Conference on Educational Multimedia and Hypermedia, Calgary, Canada. <http://taz.cs.ubc.ca/egems/reports/kamran5.doc>.
- Seely-Brown, J., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educ. Res.* 18, 32–42.
- Shayer, M., and Adey, P. (1993). Accelerating the development of formal thinking in middle and high school students. IV: Three years after a two-year intervention. *J. Res. Sci. Teach.* 30, 351–366.
- Staver, J., and Halsted, D., (1985). The effects of reasoning, use of models, sex type, and their interactions on posttest achievement in chemical bonding after constant instruction. *J. Res. Sci. Teach.* 22, 437–447.
- Stevens, R.H. (1991). Search path mapping: A versatile approach for visualizing problem-solving behavior. *Acad. Med.* 66(9), S72–S75.
- Stevens, R.H., and Najafi, K. (1993). Artificial neural networks as adjuncts for assessing medical students' problem-solving performances on computer-based simulations. *Comput. Biomed. Res.* 26(2), 172–187.
- Stevens, R.H., Kwak, A.R., and McCoy, J.M. (1992). Solving the problem of how medical students solve problems. *M.D. Comput.* 8(1), 1320.
- Stevens, R., Wang, P., and Lopo, A. (1996). Artificial neural networks can distinguish novice and expert strategies during complex problem-solving. *J. Am. Med. Inform. Assoc.* 3(2), 131–138.
- Stevens, R.H., Ikeda, J., Casillas, A., Palacio-Cayetano, J., and Clyman, S. (1999). Artificial neural network-based performance assessments. *Comput. Hum. Behav.* 15, 295–314.
- Stevens, R.H., Sprang, M., Simpson, E., Vendlinski, T., Palacio-Cayetano, J., and Paek, P. (2001). Tracing the development, transfer and retention of problem-solving skills. American Educational Research Association Symposium, Seattle, WA, presentation 44.38.
- Swiller, J., and Chandler, P. (1994). Why some material is difficult to learn. *Cognit. Instruct.* 12, 185–233.
- Underdahl, J., Palacio-Cayetano, J., and Stevens, R. (2001). Practice makes perfect: Assessing and enhancing knowledge and problem-solving skills with IMMEX software. *Learn. Lead. Technol.* 28, 26–31.
- VanLehn, K. (1996). Cognitive skill acquisition. *Annu. Rev. Psychol.* 47, 513–539.
- Vendlinski, T., and Stevens, R. (2000). The use of artificial neural nets (ANN) to help evaluate student problem solving strategies. In: *International Conference of the Learning Sciences*, University of Michigan, Mahwah, NJ: Lawrence Erlbaum Associates.
- Vendlinski, T., and Stevens, R. (2002). A Markov model analysis of problem-solving progress and transfer. *J. Technol. Learn. Assess.* 1(3), 1–20.
- Von Secker, C.E., and Lissitz, R.W. (1999). Estimating the impact of instructional practices on student achievement in science. *J. Res. Sci. Teach.* 36, 1110–1126.