

Assessing Students' Problem Solving Ability and Cognitive Regulation with Learning Trajectories

Ron Stevens, Carole R. Beal, and Marcia Sprang

Abstract

Learning trajectories have been developed for 1650 students who solved a series of online chemistry problem solving simulations using quantitative measures of the efficiency and the effectiveness of their problem solving approaches. These analyses showed that the poorer problem solvers, as determined by item response theory analysis, were modifying their strategic efficiency as rapidly as the better students, but did not converge on effective outcomes. This trend was also observed at the classroom level with the more successful classes simultaneously improving both their problem solving efficiency and effectiveness. A strong teacher effect was observed, with multiple classes of the same teacher showing consistently high or low problem solving performance.

The analytic approach was then used to better understand how interventions designed to improve problem solving exerted their effects. Placing students in collaborative groups increased both the efficiency and effectiveness of the problem solving process, while providing pedagogical text messages increased problem solving effectiveness, but at the expense of problem solving efficiency.

R. Stevens (✉)

UCLA IMMEX Project, Brain Research Institute,
UCLA School of Medicine, 5601 W. Slauson Ave. #272,
Culver City, CA 90230, USA
e-mail: immex_ron@hotmail.com

C.R. Beal

School of Information Science, University of Arizona,
P.O. Box 210077, Tucson, AZ 85721, USA

M. Sprang

Placentia-Yorba Linda Unified School District,
1830 N. Kellogg Drive, Anaheim, CA 92807, USA

Context

We have been developing reporting systems for problem solving which are helping to measure how strategically students are thinking about scientific problems and whether interventions to improve this learning are having the desired effect. The system is termed IMMEX (Interactive MultiMedia Exercises), and is an online library of problem solving science simulations coupled with layers of probabilistic tools for assessing

students' problem solving performance, progress, and retention (Soller & Stevens, 2007; Stevens & Palacio-Cayetano, 2003; Stevens, Soller, Cooper, & Sprang, 2004; Stevens, Wang, & Lopo, 1996; Cooper, Cox, Nammouz, Case, & Stevens, 2008; Thadani, Stevens, & Tao, 2009).

IMMEX problems are what Frederiksen (1984) referred to as "structured problems requiring productive thinking," meaning they can be solved through multiple approaches, and students cannot rely on known algorithms to decide which resources are relevant and how the resources should be used. IMMEX problems are rich in cognitive experiences with over 90% of the utterances of students when solving a series of cases being cognitive or metacognitive in nature (Chung et al., 2002), and is an environment where instruction can be varied and the effects of different interventions tested.

IMMEX supports detailed assessments of students' overall problem solving effectiveness and efficiency by combining solution frequencies (or IRT estimates) which are outcome measures and artificial neural network (ANN) and hidden Markov modeling (HMM) performance classifications which provide a strategic dimension (Stevens, 2007; Stevens & Thadani, 2007; Stevens & Casillas, 2006). To simplify reporting and to make the models more accessible for teachers, these layers of data can be combined into an economics-derived approach which considers students' problem solving decisions in terms of the resources available (what information can be gained) and the costs of obtaining the information.

Extensive prior research has shown that students vary widely in how systematically and effectively they approach IMMEX problems (Stevens et al., 2004; Soller & Stevens, 2007). Some students carefully and systematically look for information sources that are appropriate for the current case, keep track of the information that they are accessing, and answer when the information they have reviewed is sufficient to support the answer, whereas other students are less systematic, often reinspecting information they have already viewed (Stevens & Thadani, 2007; Soller & Stevens, 2007). In this regard, IMMEX performances are reflections of students' ability (i.e., effectiveness) as well as their regulation of cognition (i.e., efficiency).

Students who review all available problem resources are not being very efficient, although they might eventually find enough information to arrive at the right answer. Other students might not look at enough resources to find the information required to solve the problem, i.e., they are being efficient but at the cost of being ineffective. Students demonstrating high strategic efficiency should make the most effective problem solving decisions using the fewest number of the resources available. As problem solving skills are gained this should be reflected as a process of resource reduction (i.e., higher efficiency) and improved outcomes (greater effectiveness) (Haider & Frensch, 1996).

Dissecting problem solving along these two dimensions provides an opportunity to detail how classroom practices like collaborative learning or the provision of pedagogical or metacognitive prompts can influence problem solving outcomes. Do they equally affect the efficiency and effectiveness of the problem solving process or are there differential effects? This is the framing question for this study.

Theoretical Background

Most theoretical frameworks for metacognition identify two major components: knowledge of cognition (declarative and procedural knowledge) and regulation of cognition (or executive component) (Schraw, 2001; Schraw, Brooks, & Crippen, 2005; Schraw, Crippen, & Hartley, 2006). The former is often understood as metacognitive awareness and has received considerably more attention than the regulation of cognition, which comprises the repertoire of actions in which an individual engages while performing a task. Consistent with this framework, metacognition occurs when individuals plan, monitor, and evaluate their own cognitive behavior in a learning environment or problem space (Ayersman, 1995).

Despite its importance, the study of metacognition has been slowed by the lack of simple, rapid, and automated assessment tools. Technology-based learning environments provide the foundation for a new era of integrated, learning-centered assessment systems (Quellmalz & Pellegrino, 2009). It is now becoming possible to rapidly

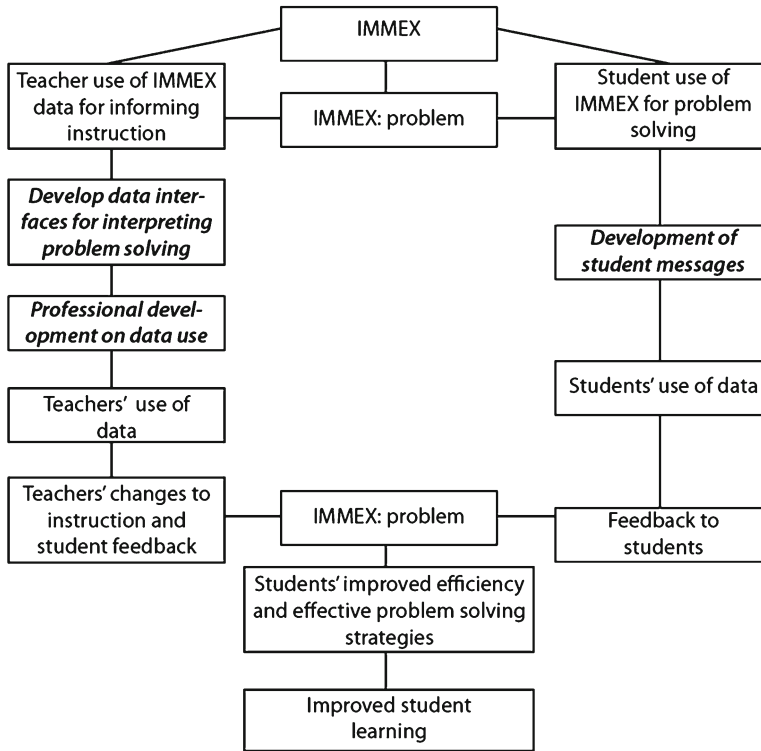


Fig. 27.1 Proposed approaches for improving student's problem solving skills

acquire data about students' changing knowledge, skill and understanding as they engage in real-world complex problem solving, and to create predictive models of their performance both within problems (Murray & VanLehn, 2000) as well as across problems and domains (Stevens et al., 2004). A range of analytic tools are being applied in these analyses including Bayesian Nets (Mislevy, Almond, Yan, & Steinberg, 1999), computer adaptive testing based on item response theory (IRT) (Linacre, 2004), and regression models and artificial neural networks (ANN) (Beal, Mitra, & Cohen, 2007; Soller & Stevens, 2007), each of which possesses particular strengths and limitations (Williamson, Mislevy, & Bejar, 2006).

How can this data be best put to use? A proposed model for improving problem solving approaches is shown in Fig. 27.1 and is based along two dimensions: (1) Teacher professional development and classroom practice and (2) direct student feedback.

Recent analyses of traditional assessment approaches and professional development models indicate that interventions often fail because teachers either do not fully understand how to implement them, or are not adequately supported in their efforts to implement them (Desimone, 2002; Lawless & Pellegrino, 2007; Spillane, Reiser, & Reimer, 2002). Simply increasing teachers' access to assessment data, however, may only exacerbate the challenges that they face in crowded classrooms when adapting instruction. Thus, new approaches are needed to provide teachers with accurate, predictive, and useful data about their students' learning in ways that are easily and rapidly understood. Data available in real time that speak to process as well as outcomes and that are intuitively easy to understand would seem to be minimum requirements.

Finding the optimum granular and temporal resolutions for reporting this assessment data will be a fundamental challenge for making the data accessible, understandable and useful for a

diverse audience (e.g., teachers, policy makers and students) as each may have different needs across these dimensions (Alberts, 2009; Loehle, 2009). If the model resolution is general and/or delayed then important dynamics of learning may be lost or disguised for teachers. If the resolution is too complex or the reporting too frequent the analysis will become intrusive and cumbersome.

Teachers however are only one side of the learning equation; we need to consider students as well. Overall, prior research suggests that students' undirected problem solving in science domains tends to be relatively unsystematic, and that students are often unselective with regard to the evidence that is collected and considered. Students' difficulties with problem solving can be especially evident in technology-based learning environments, which often require careful planning and progress monitoring to use effectively (Schauble, 1990; Stark, Mandl, Gruber, & Renkl, 1999). When students can readily explore multiple sources of information and experiment with different combinations of factors, they can easily become distracted from the primary objective of using the information to solve the problem.

One approach to improving students' problem solving is to link the technology-based activity with classroom activities designed to help students adopt good problem solving strategies and help monitor their progress. Such activities would remind students to make sure that the goal of the problem is clearly understood, identify the information that will be most helpful in solving the problem, and monitor their progress towards the solution. Adapting this approach, Schwarz and White (2005) found that students improved in their understanding of the role of models in scientific problem solving when the computer-based activity of designing models was enhanced with a classroom-based curriculum. Although the results were encouraging, one limitation was that the program was quite intensive, involving 10 weeks of classroom activities and support from university researchers. Thus, the curriculum-embedded approach might be difficult for many science teachers to implement on their own, given limited resources and constraints on classroom science activities.

Task and Analytic Approaches

The architecture of IMMEX contains a series of tasks, a student management and organization system, a data warehouse and an analytic modeling and reporting module. One IMMEX task is called *Hazmat*, which provides evidence of a student's ability to conduct qualitative chemical analyses. The problem begins with a multimedia presentation, explaining that an earthquake caused a chemical spill in the stockroom and the student's challenge is to identify the chemical. The problem space contains 22 menu items for accessing a Library of terms, the Stockroom Inventory, or for performing Physical or Chemical Testing. When the student selects a menu item, she verifies the test requested and is then shown a presentation of the test results (e.g., a precipitate forms in the liquid). Students continue to gather the information they need to identify the unknown chemical so they can solve the problem (Fig. 27.2).

Hazmat contains 38 problem cases which involve the same basic scenario but vary in difficulty due to the properties of the different unknown compounds being studied. These multiple instances provide many opportunities for students to practice their problem solving and also provide data for Item Response Theory (IRT) estimates of problem solving ability which can be useful for comparing outcomes with more traditional ability measures such as grades.

IMMEX also supports detailed analyses of students' overall problem solving effectiveness and efficiency by combining outcome measures like IRT (as a measure of overall problem solving ability), and ANN (as a measure of problem solving strategy) and hidden Markov modeling (HMM) classifications (which provide a predictive measure of problem solving progress). Sample visualizations of these formats are shown in Fig. 27.3. This layered analytical approach has been very useful from a research perspective for distinguishing gender differences in problem solving approaches (Soller & Stevens, 2007) and documenting the effects of collaborative groups during problem solving (Cooper et al., 2008).



Fig. 27.2 HAZMAT. This screen shot of *Hazmat* shows the test items available (Library, Physical Tests, Chemical Tests) on the *left side* of the screen and a sample test result of a conductivity reaction in the *center*

We have combined the measures shown in Fig. 27.3 to simplify reporting using an economics-inspired approach which considers students' problem solving decisions in terms of the resources available (what information can be gained) and the costs of obtaining the information (Stevens & Thadani, 2007).

The strategy used (or the efficiency of the approach) is described by artificial neural network analysis which is a classification system. In this system, the artificial neural network's observation (input) vectors describe sequences of individual student actions during problem solving (e.g., Run_Red_Litmus_Test, Study_Periodic_Table, Reaction_with_Silver_Nitrate). The neural network then orders its nodes according to the structure of

the data. The distance between the nodes after the reordering describes the degree of similarity between students' problem solving strategies. For example, the neural networks identified situations in which students applied ineffective strategies, such as running a large number of chemical and physical tests, or not consulting the glossaries and background information.

The neural networks also identified effective problem solving strategies such as selecting a variety of applicable tests while also consulting background information. This method is able to identify other domain-specific problem-specific strategies such as repeatedly selecting specific tests (e.g., flame or litmus tests) when presented with compounds involving hydroxides (Stevens

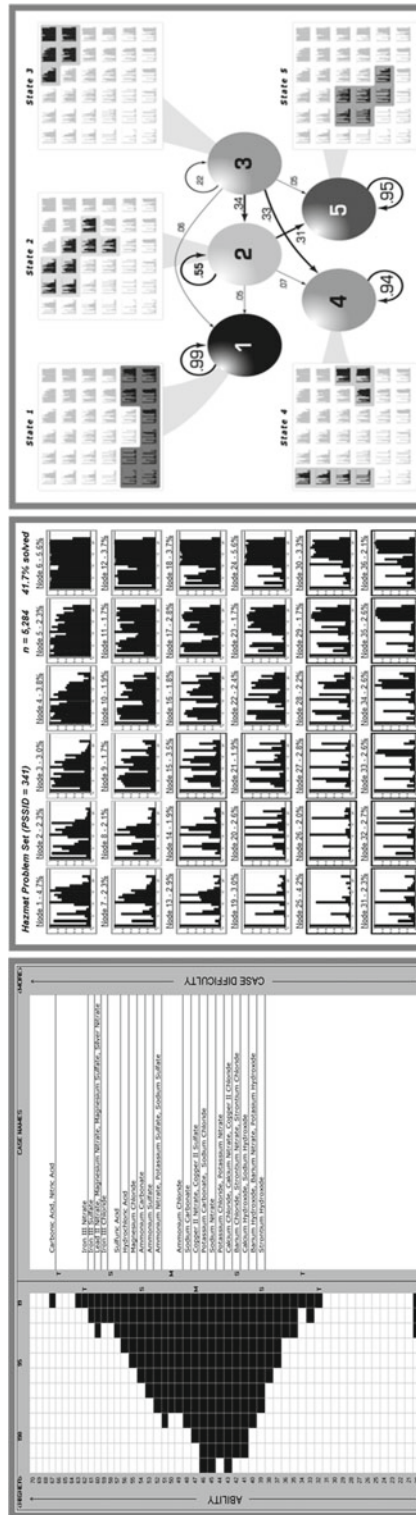


Fig. 27.3 Sample representations of item response theory ability estimates (*left*), Artificial neural network performance classifications (*middle*), and Hidden Markov modeling prediction models (*right*)

et al., 2004). Figure 27.4 (left) shows one ANN node in a 36-node network that was constructed from 5,284 performances of university and high school chemistry students. Figure 27.4 (right) shows the entire 36-node network representing the 36 different problem solving strategies used by the students. Each node of the network is represented by a histogram showing the frequency of items selected by students. For example, there were 22 tests related to Background Information (items 2–9), Flame Tests, Solubility and Conductivity (items 9–13), Litmus tests (items 14, 15), Acid and Base Reactivity (items 16, 17), and Precipitation Reactions (items 18–22).

Student performances that were grouped together at a particular node represented problem solving strategies adopted by students who always selected the same tests (i.e., those with a frequency of 1). For instance, all Node 15 performances shown in the left-hand side of Fig. 27.5 contain the items 1 (Prologue) and 11 (Flame Test). Items 5, 6, 10, 13, 14, 15, and 18 have a selection frequency of 60–80%, meaning that any individual student performance that falls within that node would most likely contain some of those items. Items with a selection frequency of 10–30% were regarded more as background noise than significant contributors to the strategy represented by that node.

The topology of the trained neural network provides information about the variety of different strategic approaches that students apply in solving IMMEX problems. First, it is not surprising that a topology is developed based on the quantity of items that students select. The upper right hand of the map (nodes 6, 12) represents strategies where a large number of tests are being ordered, whereas the lower left contains clusters of strategies where few tests are being ordered. There are also differences that reveal the quality of information that students use to solve the problems. Nodes situated in the lower right hand corner of Fig. 27.4 (nodes 29, 30, 34, 35, 36) represent strategies in which students selected a large number of items, but no longer needed to reference the Background Information (items 2–9). The classifications developed by ANN therefore reflect how students perceive the problem space,

and are regulating their test selections in response to these perceptions.

While ANN nodal classifications provide a snapshot of what a student did on a particular performance, it would be instructionally more helpful if it were possible to automatically track and report changes in strategy over time. In order to generate a time series that could potentially be predictive of future work, a series of these performances must be grouped together and classified by another type a classifier, in our case, a hidden Markov modeling technique. Similar to the training of the artificial neural network classifier a training set of hundreds/thousands of sequences of student performances are used for training where students performed 4–10 *Hazmat* cases. This training results in HMM model classifiers which can categorize future sequences of performances.

Figure 27.5 shows the results of such training and illustrates a fundamental component of IMMEX problem solving: individuals who perform a series of these simulations stabilize with preferred strategies after 2–4 problem instances. This data shows hidden Markov models of the problem solving strategies of 1,790 students who performed seven of the *Hazmat* simulations. Many students began their problem solving with a limited (these are termed State 1 strategies) or extensive search (State 3) of the problem space. These designations arise from the association of certain ANN nodal classifications with different HMM States. With practice, these strategies decreased and they became more efficient and effective (States 4 and 5).

These characterizations help in determining which students may be guessing, failing to evaluate their processes, or randomly selecting items, i.e., issues with metacognition. Several advantages of this concurrent assessment include high automation and time efficiency, minimal susceptibility to researcher's bias, and a more naturalistic problem solving setting. As described below, this type of analysis can be further collapsed into three descriptors to identify metacognitive levels: high, intermediate, and low metacognition use for comparisons with other metacognitive metrics (Cooper et al., 2008).

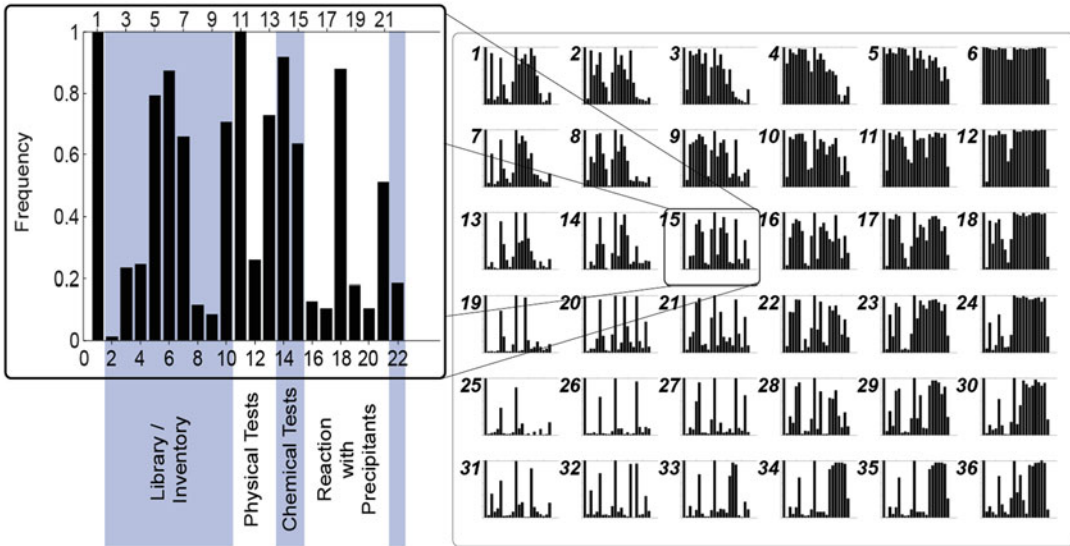


Fig. 27.4 Neural network performance patterns. The 36 Neural network nodes are represented by a 6 × 6 grid of 36 graphs. The nodes are numbered 1 through 36 left-to-right and top-to-bottom; for example, *the top row* is comprised of nodes 1 through 6. As the neural network is iteratively trained, the performances are automatically grouped into

these 36 nodes so that each node represents a different generalized problem solving strategy. These 36 classifications are observable descriptive classes that can serve as input to a test-level scoring process or linked to other measures of student achievement or cognition. They may also be used to construct immediate or delayed feedback to the student

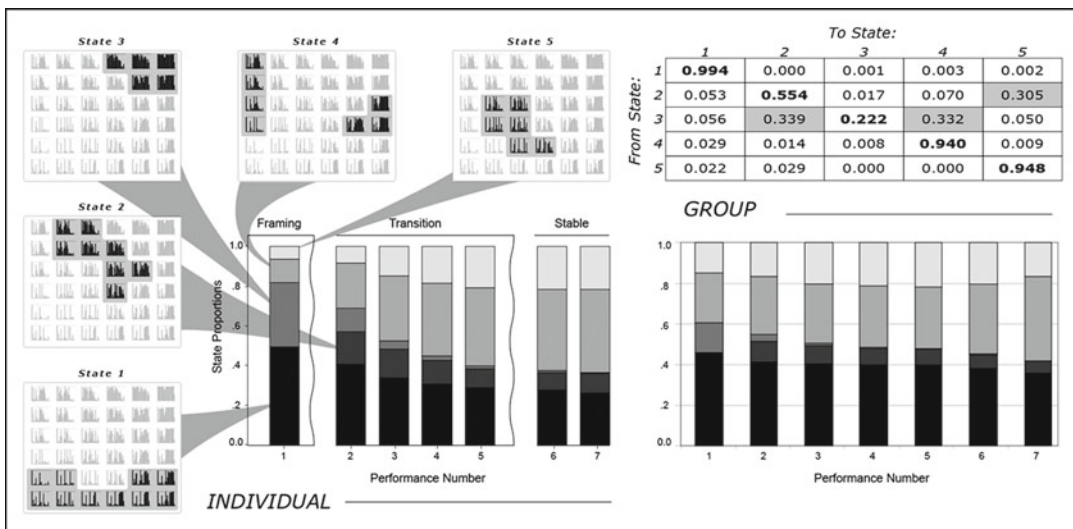


Fig. 27.5 Modeling individual and group learning trajectories. This figure illustrates the strategic changes as individual students or groups of students gain experience in *Hazmat* problem solving. Each *stacked bar* shows the distribution of HMM states for the students ($N=1,790$) after a series (1–7) of performances. These states are also mapped back to the 6 × 6 matrices which represent 36 different strategy groups identified by self-organizing ANN.

The *highlighted boxes* in each neural network map indicate which strategies are most frequently associated with each State. From the values showing high cyclic probabilities along the diagonal of the HMM transition matrix (*upper right*), States 1, 4, and 5 appear stable, suggesting once adopted, they are continually used. In contrast, students adopting State 2 and 3 strategies are more likely to adopt other strategies (*gray boxes*)

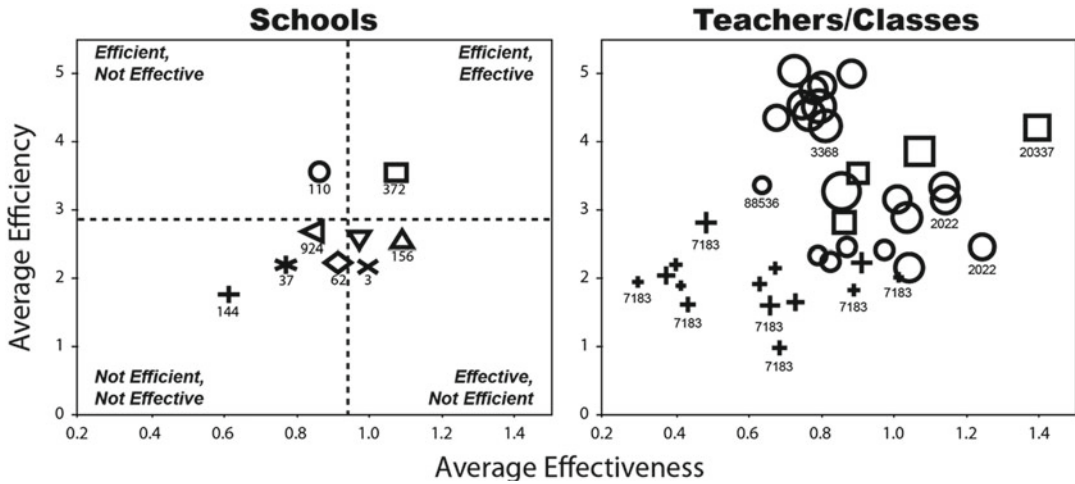


Fig. 27.6 Aggregated efficiency and effectiveness measures of schools and classrooms that performed *Hazmat*. The dataset was aggregated by schools (*left*) and then by teachers (*symbols and text*) and classrooms (*right*) and the efficiency (on a scale of 0–6) and effectiveness (on a scale of 0–2) mea-

asures calculated as described earlier. The symbol sizes are proportional to the number of performances. Each axis in (a) is bisected by *dotted lines* indicating the average efficiency and effectiveness measures of the dataset creating quadrant combinations of high and low efficiency and effectiveness

Figure 27.5 also illustrates how modifications to instruction can shift the dynamics of repetitive problem solving. The series of histograms in the right of this figure show that students in collaborative groups stabilize their strategies more rapidly than individuals and there are fewer performances where extensive searching occurs (i.e., State 3 strategies).

Learning Trajectories and Effects of Metacognition-Linked Interventions

The data gathered as students work with IMMEX provide rich, real-time assessment information along the efficiency and effectiveness dimensions. Figure 27.6 shows a modeling across schools and teachers/classrooms (66 classrooms, 62,774 performances) where an index of strategic efficiency is plotted against an effectiveness (i.e., solution frequency) rate. The quadrants generated by intersections of the averages of these measures reflect (1) mostly guessing (upper left corner), (2) performances where students view many resources, but miss the solution (lower left), (3) performances where many resources are

being viewed and the problem is being solved (lower right) and (4) the performances where few resources are used and the problem is solved (upper right). As expected by the visualization format, schools are distributed across the quadrants (Fig. 27.6, left). A second level of analysis showing problem solving performance across five teachers as well as their classrooms where the different classes of the same teacher are shown by the symbols, and the different teachers identified by numbers (Fig. 27.6, right). The clustering of the different classrooms of the teachers (for instance, the +’s in the lower left hand corner and the squares in the upper right corner), illustrates a significant teacher effect perhaps reflecting different instructional methods (Zimmerman, 2007). Follow-up classroom observation studies by Thadani et al., (2009) suggest that the teacher’s mental model of the problem space, and approach for solving the problem, can have a major effect on the approach adopted by the students.

Tracking problem solving efficiency and effectiveness as multiple *Hazmat* problems are performed creates a learning trajectory (Fig. 27.7) which is an important formative assessment tool

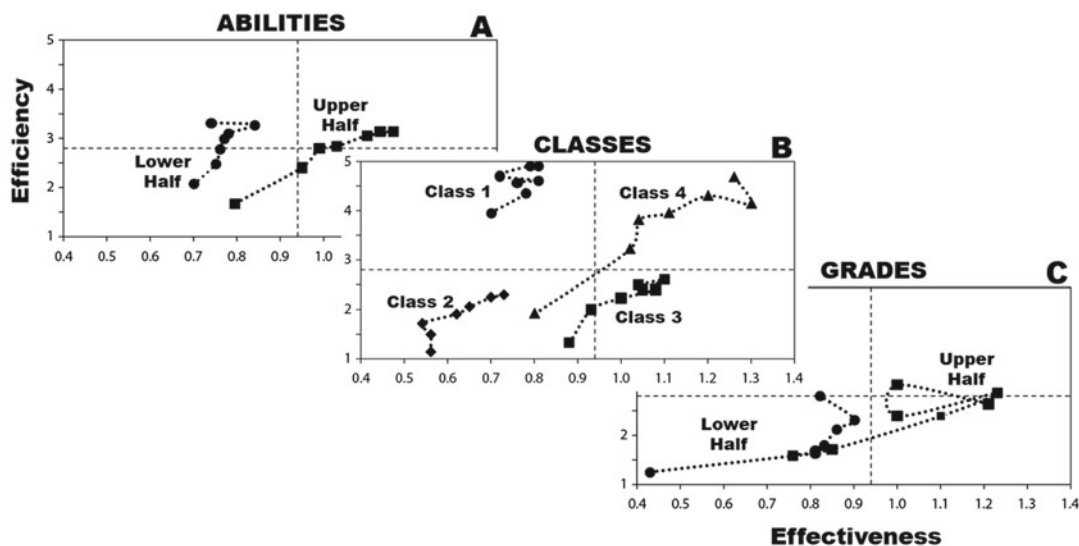


Fig. 27.7 Learning trajectories of classes and students of different abilities. (a) The dataset ($n=62,774$) was divided into lower (IRT scores=3.4–49.3) and higher (IRT scores=49.4–60.3) *Hazmat* problem solving ability students and the learning trajectories plotted. (b) The Efficiency/Effectiveness measures are stepwise plotted

for seven *Hazmat* performances for four representative classes. (c) A dataset (82 students, 780 *Hazmat* performances) for three Advanced Placement Chemistry classes was divided into high and low categories based on the final course grade and the learning trajectories calculated

showing how students improve with practice (Lajoie, 2003). Learning trajectories show that the poorer problem solvers, as determined by IRT analysis, are modifying their strategic efficiency as rapidly as the better students, as shown by the position changes along the Efficiency axis, but they are not converging on effective outcomes (Fig. 27.7a). Figure 27.7b shows that this trend can be observed in classrooms as well, (e.g., Class 1). While the more successful classes (e.g., Class 4) simultaneously improved both their problem solving efficiency and effectiveness, the lower performing classes showed gains only in efficiency. The learning trajectories are also important as changes in problem solving progress can be detected after as few as two to four student performances providing an opportunity for intervention before poor approaches have been learned. For instance, a teacher could initiate an intervention with a smaller group of students and after they have performed part of their assignment the teacher can observe online whether this was having a positive, negative or neutral effect and either continue or modify the intervention.

A similar analysis was conducted for 80 students in three Advanced Placement Chemistry classes who were separated into the upper and lower halves based on their final course grades. Again, the learning trajectories of the lower half of the students showed similar increases in strategic efficiency as the upper half of the students, but remained lower in effectiveness. (The correlations between the final grades and the efficiency index, ability estimates by IRT, and the solved rates (i.e., effectiveness) were $R^2=0.06$, $p=0.02$, $R^2=0.006$, $p=0.49$, $R^2=0.02$, $p=0.23$).

Thus from the perspectives of problem solving abilities, course grades, and perhaps the instructional environment it would appear that some students are differentially struggling with the efficiency versus effectiveness aspects of problem solving a that interventions designed to improve these skills may be useful; the question is, which intervention will work with which efficiency/effectiveness dimension? From a formative assessment perspective learning trajectories can provide evidence as to whether interventions adopted to improve learning are working.

One such approach is to integrate guidance about problem solving directly into the technology-based learning environment. Such guidance may include the types of suggestions and prompts about the metacognitive aspects of good problem solving that have been associated with effective teacher implementation and skilled instruction from expert human tutors. More specifically, good problem solvers do more than apply known procedures to familiar problems. Rather, they consider carefully the nature of the problem before starting to work, plan an appropriate approach, implement the plan, and continually evaluate progress towards the solution (Cooper & Sandi-Urena, 2009; Swanson, 1990). Good problem solvers also recognize that difficult problems may require time and effort to solve, and that some “moments in the dark” are to be expected during the problem solving process. If the kinds of metacognitive guidance provided by skilled teachers could be integrated directly into simulation learning environments, then we might expect to find students adopting better strategies.

The benefits of individualized instruction have been well documented in studies of expert human tutors, in terms of enhanced learning outcomes for novices (Lepper, Woolverton, Mumme, & Gurtner, 1993). The benefits of individualized instruction have also been documented in the context of Intelligent Tutoring Systems (ITS) software for mathematics instruction (Anderson, Carter, & Koedinger, 2000; Heffernan & Koedinger, 2002; Koedinger, Corbett, Ritter, & Shapiro, 2000). Moreno and Duran (2004) found that students who received guidance while working in a discovery-based simulation showed stronger posttest performance and higher transfer rates than students who did not receive guidance. Studies of ITS have also indicated that students who seek out and use multimedia resources show stronger learning outcomes than students who do not use the instructional resources (Wallis, Beal, Arroyo, & Woolf, 2005). While in the past ITS have primarily targeted the cognitive aspects of the student, they are increasingly being expanded to contribute to the learners' intrinsic motivation (Conati & Zhao, 2004).

Within the development and study of student feedback, we wanted to find empirical evidence of how students use direct feedback from IMMEX to help them improve the way they problem solve.

The opposite pole to individual learning is collaborative learning. As tasks have become more complex and distributed, organizations have increasingly turned to the use of teams to share the effort and most have largely become team based. It is not surprising therefore that mastering teamwork is regarded as a cornerstone of twenty-first century learning and finding ways to improve communication and collaboration is an important area of research (Partnership for 21st Century Skills, 2013). Researchers have collected evidence of metacognition development during collaborative work and through the practice of collective metacognitive activities (Case, Gunstone, & Lewis, 2001; Georgiades, 2006). Hausmann, Chi, and Roy (2004) have extensively studied the benefits that are associated with collaboration. Learning in dyads therefore would also seem like a useful potential intervention for measuring its' effects on problem solving efficiency and effectiveness.

The learning trajectory for students ($N=50,062$ performances, dotted line with open circle) who improved at their own pace is characterized by progressive improvement across both the efficiency and effectiveness dimensions which begins to plateau after around four performances (Fig. 27.8). This plateau mirrors the stabilization of strategies and abilities we have previously documented using HMM and IRT (Stevens & Casillas, 2006; Stevens & Thadani, 2007).

A second learning trajectory is from students who received text messages that were integrated into the prologue of each problem, i.e., before the student began actually working on the problem. ($n=11,497$ performances, dotted line with open square). They were specifically designed to encourage students to reflect on their problem solving. The messages appeared during the Prologue of each *Hazmat* problem (i.e., during problem framing) and were randomly selected for each case from the message bank, with the restriction that a particular message would only

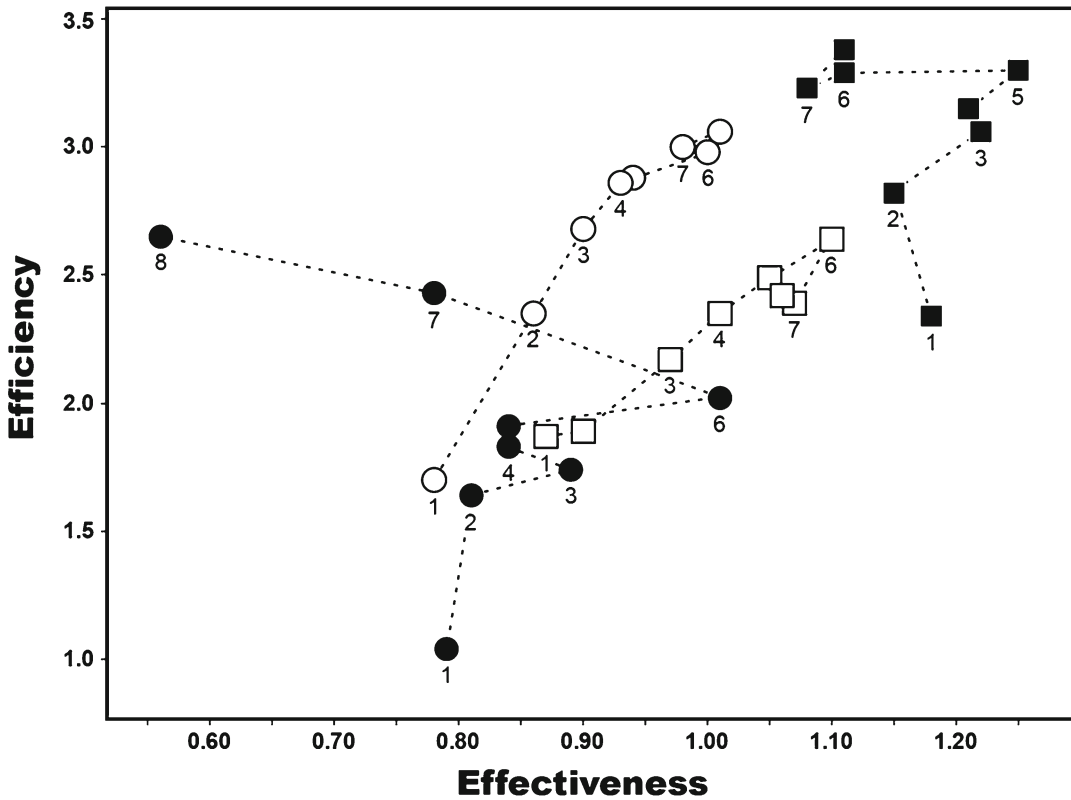


Fig. 27.8 Hazmat learning trajectories. The vertices of effectiveness and efficiency were calculated for students in different intervention groups after each of eight (*sequentially numbered*) Hazmat problem performances

be shown once to an individual student. The messages suggested for example are as follows: “When you read the IMMEX problem, don’t let yourself rush into trying different things. Stop and think for a minute first.” What have you learned in science class that could help you identify the right place to start?

It is important to note that the scaffolding messages did not provide information about the science content that would help the student solve the problem. In fact, all the relevant science content information is already available in the case; the student’s task is to think about which information might be most useful, that is, to be focused and selective. The scaffolding messages were designed to address problem solving as a process and to encourage students to focus on their actions and the goal of solving the problem (i.e., regulation), rather than to explore the sim-

ulation. Students who received the metacognitive—directed hints became less efficient, meaning that they looked at more problem materials, but they also became more effective problem solvers.

A control group of students ($n=1,215$ performances, dotted line with filled circle) also received messages during the Prologue, but here the messages were designed to be generic academic advice (e.g., “It’s a good idea to keep up with the reading for your science class.”). These students became less efficient as well as less effective. Thus, the message content was critical to improving students’ problem solving; the presence of text messages alone was not helpful. Finally, grouping students into pairs ($n=5,577$ performances, dotted line with filled square), improved both the efficiency as well as the effectiveness of the problem solving strategies.

Discussion

The studies described have traced the changes in students' problem solving ability (i.e., effectiveness) as well as their regulation of their cognition (i.e., efficiency) as they gained problem solving experience. They also showed the differential effects of interventions targeted to groups or individuals on these two problem solving dimensions. The greatest positive effect on both efficiency and effectiveness was gained by having students perform simulations in groups. In a separate study, Case et al. (2007) have shown that these positive benefits persisted when students were subsequently asked to solve additional problems on their own.

More recently Sandi-Urena et al. (2010) have shown that a non-related form of collaborative learning was sufficient to promote improved problem solving ability. Their intervention used a pretest/posttest experimental design. The intervention was a three phase "problem solving" activity that involved neither a chemistry problem nor was it directly associated with the IMMEX assessment system or problem solving activities. The intervention took place over 3 weeks. Phase one involved a small group collaborative problem solving activity and was designed to promote metacognition by the use of prompts and social interaction. The problems required students to sort through extraneous information and could not be solved by rote methods or without monitoring and evaluating their progress (core components of metacognitive skillfulness). Phase two, where students solved another problem for homework, was designed to promote individual reflection, and phase three provided students with feedback and summaries of their activities. Students were asked to reflect on what they had learned during the process and what it meant for their approach to future problem solving activities.

A comparison of student performances before and after this intervention indicated that they used more efficient strategies, and had higher problem solving ability after the intervention. Even though there was no explicit link between

the metacognitive intervention and the IMMEX problems, the intervention made students more likely to monitor and evaluate their progress through the problem, leading to increased problem solving ability.

The interventions targeted to individuals also shifted the shapes of learning trajectories. The inclusion of pedagogical messages or hints while the students were framing the problem showed different effects depending on the content of the messages. The messages that were designed with metacognition in mind improved the ability of the student to solve problems, but decreased the efficiency of the process, e.g., they seemed to make the students more reflective or cautious. This was, in fact the goal of these messages, to foster improved cognitive regulation. The messages that were general study aids also had an effect on the students' problem solving in that they decreased both the efficiency and the effectiveness of the problem solving, i.e., they were deleterious along both dimensions. While the possibility exists that they may have been a problem solving distraction for the students, given the magnitude of the effects we chose not to include such messages in subsequent studies.

Recently these studies have been extended to middle school classrooms using an IMMEX problem set called *Duck Run* (Beal & Stevens, 2011). This is also a chemistry problem set where the prologue describes that an unknown substance has been illegally dumped into a local duck pond, possibly putting the local wildlife at risk. The student's task is to identify the substance so that it can be properly removed. Students who worked with the message-enhanced version were more likely to solve the problems and to use more effective problem solving strategies than students who worked with the original version. Benefits of the messages were observed for students with relatively poor problem solving skills, and for students who used exhaustive strategies. It would seem therefore that the beneficial effects of well-constructed messages immediately prior to problem solving are generalizable to multiple grade levels.

Combined these studies show that technology can provide dynamic models of what students are doing as they learn problem solving without

creating a burden on educational systems. While illustrated for chemistry, such models are applicable to other problem solving systems where learning progress is tracked longitudinally. When shared with teachers and students in real time they can provide a roadmap for better instruction by highlighting problem solving processes and progress and documenting the effects of classroom interventions and instructional modifications. The differences observed across schools, teachers, and student abilities shifts the focus to the classroom and may provide a means for matching students and instruction or matching teachers with professional development activities.

Acknowledgments Supported in part by National Science Foundation Grants DUE 0512203 and ROLE 0528840 and by a grant from the US Department of Education's Institute of Education Sciences (R305H050052).

References

- Alberts, B. (2009). Redefining science education. *Science*, 323(5913), 437.
- Anderson, J. R., Carter, C., & Koedinger, K. R. (2000). Tracking the course of mathematics problems. National Science Foundation, ROLE Award, Carnegie Mellon University.
- Ayersman, D. J. (1995). Effects of knowledge representation format and hypermedia instruction on metacognitive accuracy. *Computers in Human Behavior*, 11(3–4), 533–555.
- Beal, C. R., Mitra, S., & Cohen, P. R. (2007). Modeling learning patterns of students with a tutoring system using Hidden Markov Models. *Proceedings of the 13th International Conference on Artificial Intelligence in Education*. Amsterdam: IOS press.
- Beal, C. R., & Stevens, R. (2011). Improving students' problem solving in a web-based chemistry simulation through embedded metacognitive messages. *Technology Instruction, Cognition and Learning*, 8(3–4) 255–271.
- Case, J., Gunstone, R., & Lewis, A. (2001). Students' metacognitive development in an innovative second year chemical engineering course. *Research in Science Education*, 31, 313–335.
- Case, E., Stevens, R., & Cooper, M. M. (2007). Is collaborative grouping an effective instructional strategy? Using IMMEX to find new answers to an old question. *Journal of College Science Teaching*, 36(6), 42.
- Chung, G. K. W. K., deVries, L. F., Cheak, A. M., Stevens, R. H., & Bewley, W. L. (2002). Cognitive process validation of an online problem solving assessment. *Computers and Human Behavior*, 18(6), 669–684.
- Conati, C., & Zhao, X. (2004). Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game. *Proceedings of the 9th International Conference on Intelligent User Interfaces* (pp. 6–13). Funchal, Madeira, Portugal.
- Cooper, M. M., Cox, C. T., Jr., Nammouz, M., Case, E., & Stevens, R. H. (2008). An assessment of the effect of collaborative groups on students' problem solving strategies and abilities. *Journal of Chemical Education*, 85(6), 866–872.
- Cooper, M. M., & Sandi-Urena, S. (2009). Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving. *Journal of Chemical Education*, 86(2), 240–245.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479.
- Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. *Review of Educational Research*, 54(3), 363–407.
- Georghiades, P. (2006). The role of metacognitive activities in the contextual use of primary pupils' conceptions of science. *Research in Science Education*, 36, 29–49. doi:10.1007/s11165-004-3954-8.
- Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, 30(3), 304–337.
- Hausmann, R. G., Chi, M. T. H., & Roy, M. (2004). Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. *26th Annual Conference of the Cognitive Science Society* (pp. 547–552). Chicago, IL.
- Heffernan, N. T., & Koedinger, K. R. (2002). An intelligent tutoring system incorporating a model of an experienced human tutor. *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems*, Biarritz, France.
- Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. J. (2000). *Carnegie learning's cognitive tutor: Summary research results. White paper*. Pittsburgh: Carnegie Learning.
- Lajoie, S. P. (2003). Transitions and trajectories for studies of expertise. *Educational Researcher*, 32(8), 21–25.
- Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: knowns, unknowns, and ways to pursue better questions and answers. *AERA Review of Educational Research*, 77(4), 575–614.
- Lepper, M. R., Woolverton, M., Mumme, D., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 75–105). Hillsdale: Erlbaum.
- Linacre, J. M. (2004). *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com.
- Loehle, C. (2009). A guide to increased creativity in research—Inspiration or perspiration. *Bioscience*, 40, 123–129.
- Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (1999). Bayes nets in educational assessment: Where

- do the numbers come from? In K. B. Laskey & H. Prade (Eds.), *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 437–446). San Francisco: Morgan Kaufmann.
- Moreno, R., & Duran, R. (2004). Do multiple representations need explanations? The role of verbal guidance and individual differences in multimedia mathematics learning. *Journal of Educational Psychology, 96*, 492–503.
- Murray, R. C. & VanLehn, K. (2000). A decision-theoretic, dynamic approach for optimal selection of tutorial actions. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent Tutoring Systems, Fifth International Conference, ITS 2000, Montreal, Canada* (pp. 153–162). New York: Springer.
- Partnership for 21st Century Skills. Retrieved February 8, 2013, <http://www.p21.org>.
- Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science, 323*(5910), 75–79.
- Sandi-Urena, S., Cooper, M. M., & Stevens, R. H. (2010). Enhancement of metacognition use and awareness by means of a collaborative intervention. *International Journal of Science Education*. doi:10.1080/09500690903452922. Retrieved from <http://dx.doi.org/10.1080/09500690903452922>. First published on: 2 February 2010 (iFirst).
- Schauble, L. (1990). Belief revision in children. *Journal of Experimental Child Psychology, 49*, 31–57.
- Schraw, G. (2001). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 3–16). Dordrecht: Kluwer.
- Schraw, G., Brooks, D. W., & Crippen, K. J. (2005). Using an interactive, compensatory model of learning to improve chemistry teaching. *Journal of Chemical Education, 82*(4), 637–640.
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111–139.
- Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction, 23*(2), 165–205.
- Soller, A., & Stevens, R. H. (2007). Applications of stochastic analyses for collaborative learning and cognitive assessment. In G. Hancock & K. Samuelson (Eds.), *Advances in latent variable mixture models*. Charlotte: Information Age Publishing.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and re-focusing implementation research. *Review of Educational Research, 72*(3), 387–431.
- Stark, R., Mandl, H., Gruber, H., & Renkl, A. (1999). Instructional means to overcome transfer problems in the domain of economics: Empirical studies. *International Journal of Educational Research, 31*, 591–609.
- Stevens, R. H., & Palacio-Cayetano, J. (2003). Design and performance frameworks for constructing problem-solving simulations. *Cell Biology Education, 2*(3), 162–179.
- Stevens, R. H., Soller, A., Cooper, M., & Sprang, M. (2004). Modeling the development of problem solving skills in chemistry with a web-based tutor. In J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds.), *Intelligent Tutoring Systems, 7th International Conference Proceedings* (pp. 580–591). Berlin: Springer.
- Stevens, R. H., Wang, P., & Lopo, A. (1996). Artificial neural networks can distinguish novice and expert strategies during complex problem solving. *Journal of the American Medical Informatics Association, 3*(2), 131–138.
- Stevens, R. H. (2007). A value-based approach for quantifying student's scientific problem solving efficiency and effectiveness within and across educational systems. In R. W. Lissitz (Ed.), *Assessing and modeling cognitive development in school* (pp. 217–240). Maple Grove: JAM.
- Stevens, R. H., & Thadani, V. (2007). Quantifying student's scientific problem solving efficiency and effectiveness. *Technology, Instruction, Cognition and Learning, 5*(2–3–4), 325–337.
- Stevens, R. H., & Casillas, A. (2006). Artificial neural networks. In R. E. Mislav, D. M. Williamson, & I. Bejar (Eds.), *Automated scoring of complex tasks in computer based testing: An introduction* (pp. 259–312). Mahwah: Lawrence Erlbaum.
- Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82*(2), 306–314.
- Thadani, V., Stevens, R. H., & Tao, A. (2009). Measuring complex features of science instruction: developing tools to investigate the link between teaching and learning. *Journal of the Learning Sciences, 18*(2), 285–322.
- Wallis, R., Beal, C. R., Arroyo, I., & Woolf, B. P. (2005, April). Cognitive predictors of response to web-based tutoring. Accepted for presentation at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.
- Williamson, D. M., Mislav, R. J., & Bejar, I. I. (Eds.). (2006). *Automated scoring of complex tasks in computer based testing*. Mahwah: Erlbaum Associates.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172–223.

Part V

Multi-agent Systems to Measure and Foster Metacognition and Self- Regulated Learning

Using Trace Data to Examine the Complex Roles of Cognitive, Metacognitive, and Emotional Self-Regulatory Processes During Learning with Multi-agent Systems

Roger Azevedo, Jason Harley, Gregory Trevors,
Melissa Duffy, Reza Feyzi-Behnagh,
François Bouchet, and Ronald Landis

Abstract

This chapter emphasizes the importance of using multi-channel trace data to examine the complex roles of cognitive, affective, and metacognitive (CAM) self-regulatory processes deployed by students during learning with multi-agent systems. We argue that tracing these processes as they unfold in real-time is key to understanding how they contribute both individually and together to learning and problem solving. In this chapter we describe MetaTutor (a multi-agent, intelligent hypermedia system) and how it can be used to facilitate learning of complex biological topics and as a research tool to examine the role of CAM processes used by learners. Following a description of the theoretical perspective and underlying assumptions of self-regulated learning (SRL) as an event, we provide empirical evidence from five different trace data, including concurrent think-alouds, eye-tracking, note taking and drawing, log-files, and facial recognition, to exemplify how these diverse sources of data help understand the complexity of CAM processes and their relation to learning. Lastly, we provide implications for future research of advanced learning technologies (ALTs) that focus on examining the role of CAM processes during SRL with these powerful, yet challenging, technological environments.

R. Azevedo, Ph.D. (✉) • J. Harley • G. Trevors
• M. Duffy • R. Feyzi-Behnagh • F. Bouchet
Laboratory for the Study of Metacognition and Advanced
Learning Technologies, Department of Educational and
Counselling Psychology, McGill University, 3700
McTavish Street, Montreal, QC, Canada H3A 1Y2
e-mail: roger.azevedo@mcgill.ca

R. Landis
Illinois Institute of Technology, College of Psychology,
3105 S. Dearborn St., Life Sciences #252, Chicago,
IL 60616, USA
e-mail: rlandis@iit.edu

The widespread use of advanced learning technologies (ALTs) poses numerous challenges for learners of all ages. Learning with these non-linear, multi-representational, open-ended learning environments typically involves the use of numerous self-regulatory processes, such as planning, cognitive strategies, metacognitive monitoring and regulation, emotions, and motivation. Unfortunately, learners do not always monitor and regulate these processes during learning