Distributed Interactive Learning Environments

Beverly Park Woolf
Department of Computer Science, University of Massachusetts, Amherst, Massachusetts, U.S.A.

bev@cs.umass.edu

Abstract: This chapter describes four systems that introduce new pedagogy, address inequities in the classroom and show improved learning with reduced cost. The first system supports inquiry and case-based learning, moving students towards more active learning, engagement, and hypothesis construction. The next system describes mathematics and geometry tutors that positively influence a student’s confidence and image of mathematics ability. The final system is a web-based homework system used by more than 50,000 students in 20 domains, including physics, chemistry and geology is also discussed. These systems provide solid evidence that electronic teaching can result in increased learning, reduced costs, and improved motivation.

1 Introduction

This chapter describes four learning environments that reason about individual students, model the knowledge to be taught, provide assessment of learning and customize teaching based on student needs. In these systems, students are engaged in active and authentic learning; they are not treated as blank slates with respect to goals, opinions, knowledge and time.

Issues of pedagogy and style need to be addressed before widespread successful Web learning becomes available and appropriate on a global scale. Knowledge is the Web’s greatest potential advantage and yet web resources are not automatically knowledge-based. Typical Web materials often provide unstructured, static and un-prioritized information. Pages and portals that deliver a linear selection of topics, mimicking classroom slides or notes, are difficult for students to absorb. Knowledge-centered means that the environment teach knowledge and skills based on an analysis of student needs and what students will be able to do when they finish with the materials. Web material should also track students and teach to their learning needs. Pages of slides, text, virtual laboratories and simulations are not learner-centered if they present the same material to all students. The systems described here consider the learner’s preconceptions, needs, strengths and interests. The final requirement is to ensure that a student’s thinking is visible so she can have multiple chances to revise and learn about her own learning. Assessment also involves providing data to teachers so they can assess the effectiveness of their teaching materials and possibly modify strategies based on input about student knowledge. All these system do provide this level of assessment.

The first system, an inquiry tutor is focused on changing classroom learning, by supporting a student to engage in critical thinking, to propose a hypothesis and find data that supports or refuses that hypothesis, Section 2. The second system uses multimedia to teach geometry
concepts and customizes hints for individual students, Section 3. The third system, which teaches arithmetic, identifies a student’s cognitive development and then adapts the learning to those cognitive differences, Section 4. The last system is an infrastructure for delivery of homework problems and has been used in more than twenty domains.

2 New pedagogy: Inquiry learning

The first project we describe supports a new pedagogy that is difficult to deliver in a traditional classroom setting. This is case-based and inquiry learning. The system tracks, analyzes and then comments on the student’s selection of data and creation of hypotheses and inferences. It includes coaches to encourage students to support or refute hypotheses with sufficient evidence. The infrastructure will be applied in several science domains; we have developed prototypes in biology, forestry, civil engineering and geology (Woolf, Reid, Stillings, Bruno, Murray, Reese, Peterfreund and Rath, 2001). Coaches advise students about illogical statements and inconsistent reasoning. A coach can help students organize and qualify their knowledge in preparation for acquiring more knowledge and understands (to some extent) the reasoning behind students’ hypotheses.

Evaluations of inquiry learning suggest that working in small groups has helped students develop more sophisticated strategies for coming to decisions than would be possible if students worked alone or turned only to faculty for answers (Wenk, 1999). The software is intended to provide sufficient individual support to enable class time and text assignments to be used to introduce new content. Class sessions can be devoted to team meetings with occasional short lectures fit in as needed. Developing more effective interactive electronic software and communication among students and between students and faculty should address many issues.

Rashi, the inquiry tutor we have built, encourages students to monitor their own scientific thinking. Customized feedback based on an assessment of students’ learning, draws students into dialogues. As students’ comfort levels increase, they solve more complex cases and use their reasoning on novel cases. The software scaffolds students to synthesize observations, request data to confirm or refute hypotheses and trust the validity of their own hypotheses. It supports observation of phenomena and critical reflection.

<table>
<thead>
<tr>
<th>Challenges of Inquiry-based Teaching</th>
<th>Rashi Inquiry Software Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students need scaffolding to pose open ended and authentic questions.</td>
<td>Rashi provides multiple cases in which strategies to develop a differential diagnosis are scaffolded.</td>
</tr>
<tr>
<td>Students need to plan queries and do research. They should recognize salient data and recognize issues which require further study.</td>
<td>Students identify data that are « known » or « should be studied » from a document library, an examination tool, Figure 4, and a video interview, Figures 2-3.</td>
</tr>
<tr>
<td>Students need to be mindful of the processes they use to solve problems and to monitor their own progress.</td>
<td>Rashi tracks students’ observations, data, hypotheses and support for inferences. Students move freely throughout activities (Inquiry Notebook, Figure 5).</td>
</tr>
<tr>
<td>Students should recognize when data do and do not support their hypotheses.</td>
<td>The coach helps students organize and identify consistent data.</td>
</tr>
</tbody>
</table>

Table 1 : Challenges faced by students in inquiry learning and how Rashi addresses these issues.

Tracking student reasoning. The Rashi software supports student reasoning about medical cases (Woolf, Marshall, Mattingly, Lewis, Wright, Jellison and Murray, 2003; Woolf, Reid, Stillings, Bruno, [1]

---

[1] Rashi was a biblical scholar who introduced inquiry methods in the eleventh century. He wrote extensive commentaries, produced queries, explanations, interpretations and discussions of each phrase and verse of the bible. Rashi’s written commentary on the bible made it more comprehensible for everyday scholars. Today, these and other commentaries, assembled in the Talmud, have been extended to nearly 40 volumes and continue as a source of biblical law (Steinsaltz, 1976).
Murray, Reese, Peterfreund and Rath, 2001). Table 1 lists some of the challenges faced by students who begin to use inquiry in the classroom and how Rashi begins to meet those challenges. Facilitating inquiry in a classroom is difficult for faculty, who may not know how to encourage students to ask their own questions, refine them so they can be answered through analyzing evidence, and plan to gather such evidence in laboratory or library (Stillings, Ramirez and Wenk, 2000).

**Inquiry Cycle.** In the *Orientation and Observation Phase* students are presented with a case description, situations and goals. They are guided to observe and synthesize their observations. In the biology module, a patient is presented with symptoms, Figure 1, including fatigue, weight loss, anxiety, and sweaty palms. Students try to diagnose the cause of these symptoms. They extract pertinent information and try to recognize the difference between pure observation and inference, and both of these may be taking place. The patient’s complaints form an initial set of data from which the student begins the diagnostic process.

In the *Hypotheses Phase* students brainstorm and list predictions or subgoals that might resolve some aspect of the problem. They type in causes for the observed phenomena and later predict data that will either support or weaken the hypotheses. An initial hypothesis might be vague: « I think she sounds like she might have an anxiety disorder. » Hypotheses of this sort require further development with knowledge through resources available in the tutor. Alternatively, students might wish to act on intuition or experiential knowledge such as « I bet the patient has mono, because her symptoms sound exactly like my roommate’s ». Often new terminology must be learned and understood.

In the *Collect Data Phase* students gather and request data to confirm or refute the hypothesis and resolve open questions. For example, students interview the patient about her symptoms, (Figures 2-3), perform an examination, and request medical history, or lab tests (Figure 4). Students request data to eliminate or support their hypothesis and learn to assess evidence that bears on their hypothesis independent of a mentor’s input. Rashi provides an *Inquiry Notebook* in which the students record open questions and hypotheses (Figure 5). Data may reveal flaws in hypotheses, in which case students can revise their hypotheses, change their opinion of how strongly data support or refute hypotheses, and generate new hypotheses. In the *Communication Phase*, students type in their reports, including all the selected data, inferences and hypotheses. These are sent electronically to the teacher for evaluation. In some cases, a sequential review of all observations, hypotheses, data, and explanations is presented graphically, and can be edited and re-ordered to look for patterns. At some point each student makes a
final submission which involves designating one hypothesis as the « best ». Then they turn in the *Inquiry Notebook* complete with all the competing hypotheses and their arguments for eliminating them.

Students are expected to identify key elements of scientific problems, determine the essential information needed, seek out information, clearly and concisely write the rationale for their decisions, discuss findings with peers, and work together to synthesize the approach to a solution. Students are presented with information about a patient who has certain complaints. They organize their discussions around three questions: What do we know? What do we think we know? What more do we need to know?

As each student moves through the inquiry cycle, the tutor follows her reasoning by matching it with an expert’s system of the medical case. As an example, we describe an analysis of one student’s reasoning in this case. The student interviews the patient and records important symptoms (Figures 2-3). She « performs » a medical exam on the patient and identifies salient symptoms and lab results (Figure 4). In particular, the student isolates those data that support, refute, or have no bearing on a given hypothesis. The student reads medical source documents and studies the patient’s signs and symptoms. The computer keeps track of the student’s activities and issues prompts through the expert system if the student has failed to explore some range of information. The Rashi software responds to the student with carefully crafted questions, never revealing directly the solution to the case.

**Inquiry Notebook.** The *Inquiry Notebook* collects student information (Figure 5). Once the student is oriented to the goal of the case and uses data gathering tools and libraries and web sites and textbooks, she is asked to record meaningful units of data or propositions. The student keeps track of where propositions come from (i.e., cites the sources). Students must indicate relationships between propositions by linking them with *supports/refutes* links, and these relationships become propositions as well. Relationships are often justified by some fact or principle found in a medical source. Finally, these chains of relationships terminate in hypothesis.

Students may ask the electronic *Coach* for an assessment of their work or their argument supporting or eliminating a hypothesis. The *Coach* might also interrupt the student and offer critiques and suggestions during the process. The *Coach* analyzes the student’s *Inquiry Notebook* and history of activities and gives feedback about how best to proceed. The Bayesian Belief Network (BBN), the basis of the *Coach*’s performance, comments about the *syntactic structure* of the student’s argument (Does the student understand the difference between data and hypotheses?), and the *semantic content* of the student’s argument? (Are the inferences and conclusions made supported by the data and the medical knowledge?)

![Figure 4. Patient Examination Tool](image1)

*Figure 4. Patient Examination Tool enables students to measure weight, pulse, blood pressure, etc. In this example the student has selected the head and is given choices of viewing exam results for eyes, ears, neck, etc.*

![Figure 5. The Inquiry Notebook](image2)

*Figure 5. The Inquiry Notebook. Student observations in the exam and the interview are automatically recorded in the inquiry notebook. The student indicated type (observations, inferences and hypothesis).*
3. New technology: Multimedia in geometry

The second project utilizes new technology, specifically 3-D animation, web-delivery and student modeling, to address inequities in learning. High stakes achievement, such as the Scholastic Aptitude Test (SAT), used for entrance into many colleges, have become increasingly important in the past years in the United States, and a student’s performance on such tests can have a significant impact on his or her access to future educational opportunities. At the same time, concern is growing that the use of high stakes achievement tests simply exacerbates existing group differences, and puts female students and those from traditionally underrepresented minority groups at a disadvantage.

New approaches are required to help all students perform to the best of their ability on high stakes tests. We have created a system, called « Wayang Outpost », an Intelligent Tutoring System to prepare students for the mathematics section of the (SAT). Wayang uses multimedia on the front end and tutoring decisions on the back end to address factors that make females score badly in these tests and to effectively tutor every student. Cognitive abilities such as spatial abilities and math fact retrieval are important determinants of a student’s performance on geometry problems. Mental rotation discrepancy between boys and girls accounts for significant gender differences in mathematics SAT scores (Casey, Nuttal and Pezaris, 1997). By reasoning about the underlying cognitive abilities of each student, the system helps prepare both genders for SAT-Math, probably reducing the gender gap in such test.

Wayang Outpost is a multimedia-based intelligent tutor, with both a highly interactive front-end and sophisticated intelligent approaches for problem and help selection. Multimedia is used to provide appeal for the system and also for tutoring – the hints are animations that may include rotating triangles, lines that break apart, characters pointing to sections of the problem, drawing and gesturing while talking. An Indonesian shadow puppet (Figure 6) acts as a guide to direct attention, demonstrate processes involved in the solution of a problem, and motivate students through the incorporation of an interesting setting and appealing characters. Virtual role models based on real scientists such as Anne, an orangutan expert, (Figure 8), appear in the adventures and ask for the student’s help. The setting is a Macromedia Flash animated hut based in a research station in the Kalimantan rainforest (formerly Borneo) populated by researchers, students, and indigenous Wayang art. Students and researchers address environmental issues around the orangutan endangered species (stopping fires, illegal logging, curing orangutans), providing a rich real-world content for mathematical problems.

Each math problem is presented in a movie built in Macromedia Flash, including an animated character based on the traditional Indonesian art form of shadow puppetry (Figure 6). Two forms of help are available: 1) a fairly traditional approach, such as setting up a formula or equation; and 2) a visual approach, e.g., showing the rotation of part of a figure would provide key information. If the student answers incorrectly, or requests help, the shadow puppet guide provides step by step instruction and guidance in the form of Flash animations with audio. For example, on a geometry problem, the student may see an angle with a known value rotate and move over to a corresponding angle with an unknown value on a parallel line, thus emphasizing this concept, while the puppet-teacher describes the relation.

Wayang Outpost is a hierarchy of Flash movies played within each other in the front-end. There is an environment movie -- the village setting in Borneo, the shadow-puppet tutoring character, the hut where students focus on mastering SAT skills they will later use in the rainforest. Each problem is a Flash movie, displayed within the environment. Specific problems are brought from a central server when needed and played within a section of the screen. When a hint is to be displayed, the Flash problem
execution « jumps » to the corresponding hint label, to play the hint. The full set of hints for each problem can include dozens of movie and sound files.

Wayang Outpost provides instruction via a website, ensuring easy access to students either at home or from any school connected to the Internet. Use of the web-delivery format allows easy expansion and frequent updating of the system, without requiring users to download or reinstall the program. As the student works through a problem, performance data (e.g., latency, answer choice) are stored in a centralized database. This data constitutes the « episodic » student model, i.e. raw data about every student interaction with the system. From this data, the system will make inferences on an ongoing basis to select problems at the appropriate level of challenge, and hints that will be helpful for the student. Communication between Flash and the Java-based tutor is via XML calls.

**Modeling the student.** Most intelligent tutors select activities and hints based on students’ past performance, but rarely incorporate general characteristics of the student for their predictions. We know that student general characteristics are relevant determinants of user behavior from past research (Arroyo, Beck, Woolf, Beal and Schultz, 2000), so we incorporate several characteristics in the student model. These characteristics are students’ gender, spatial ability as measured in the Vandenberg mental rotations test (Vandenberg and Kuse, 1978), and math fact retrieval speed (Royer, Tronsky, Chan, Jackson and Merchant, 1999). The student goes through a series of standard tests before starting to use the tutoring system, tests that evaluate these cognitive abilities. For each SAT-Math problem, data from performance in the tutor will be used to identify the most critical cognitive skills predicting successful solution, which may differ by gender. Each problem includes two types of hints: one based on an algebraic, step-by-step approach, the second based on mental rotation and visual estimation (a strategy that past research suggests is more often used by male students). The choice of hint type can be customized for individual students on the basis of their cognitive profile, to help them develop strategies and approaches that may be more effective for particular problems. For example, women who score low on the spatial ability assessment might receive a high proportion of hints that emphasize mental rotation and estimation, approaches that women often avoid even though they are generally more effective in a timed testing situation.

**Short-term and long-term transfer problems.** SAT-Math consists of non-traditional multiple choice problems. Figure 7 shows an example of such a problem. The fact that problems are « non-traditional » makes them very different from each other. We decided it would be important to have at least pairs of similar problems (i.e. that involve the same skills) in order to appreciate students’ learning. We created what we call « short -term transfer » problems, which are variations of an original problem in terms of the operands in the problem and probably a rotation or flip of the figure/table/graph if any. Past research suggests that by observing improvement in subsequent similar problems, we can appreciate learning closely (Arroyo, Beck, Woolf, Beal and Schultz, 2000); Arroyo, Beal, Bergman, Lindenmuth, Marshall and Woolf (2003). Although it is important to help achieve their potential on high stakes tests, the ultimate goal for educators
is to enhance students' conceptual understanding of mathematical concepts, and their ability to draw on their skills to solve novel problems. Therefore, Wayang Outpost incorporates animated « adventures » which present the student with a challenging mathematical problem which requires multiple steps and skills to solve. These adventures are designed to incorporate real-world content appropriate to the setting (e.g., an animated character based on Dr. Anne Russon of York University, an expert on the orangutan, invites the student to rescue orangutans that are trapped in a fire in the rainforest, Figure 8). Performance on the adventure math problems becomes a measure of transfer of math skills from the multiple-choice SAT-Math tutor to a different context where the same mathematics skills are needed to solve problematic situations, such as calculating shortest paths, determining if it is safe to jump down a cliff, estimating how much gas is needed, etc. Long-term transfer adventures are used as pre and post-tests to the tutoring section. These adventures have three main advantages: 1) motivation: adventures are entertaining (almost a video game), so we don’t expect students to find them burdensome; 2) the outcome of the game pre-test will serve to initialize the student model; 3) the outcome of the post-test determines the effectiveness of tutoring.

Students are not only affected by the amount of help but also by the format of the hints (Arroyo, Beck, Woolf, Beal and Schultz, 2000). We found two very distinctive ways to solve geometric problems: one that relies heavily on visualization (e.g. mental rotations, figure decomposition) and estimation, and a second one that relies on fact retrieval and application (e.g., application of formulas, solving of equations). Based on this analysis, each problem is accompanied by hints for the two kinds of solutions: one based on an analytic approach, the second based on a spatial approach. These two strategies will affect students of different spatial abilities and math fact retrieval abilities differently.

Pedagogical decisions. The main pedagogical decisions have to do with problem and help selection. The tutor first chooses a problem that is not too easy and not too hard (within the ZPD). This means to us that the student will make a small number of mistakes, or ask for a small amount of help. A second goal is that the student responds in a certain amount of time, as speed is an important matter in these tests. One last goal for the ideal problem is that the student does not « skip » it without answering (something the system allows).

The tutor next reasons about which hints to choose. Each problem has associated a partially ordered solution plan, with a hint and a main skill associated for each of the steps. Internally, Wayang keeps a partially ordered solution plan for each problem, such as the one shown in Figure 7. Showing all hints along a path is in general « too much » help, as we expect the student will have difficulty in a subset of the steps but not all of them. The tutor has to determine at which step in the solution plan the student has most likely failed, and provide only a summary of the steps before the faulty one. Also, the tutor has to decide at each bifurcation what path to take. Later sections discuss how these decisions will be made.

Evaluation studies. Evaluation studies of Wayang Outpost will be conducted in the coming academic year. The primary measures will be performance on SAT-Math problems, in a pre and post-test design, as a function of the type of tutoring provided by the system (more spatial
versus more algebraic) and the student's cognitive profile. In addition, data on student performance as a function of hints will also be used to assess the effectiveness of the tutor's help. Specifically, a student who receives help on a problem of one type should achieve correct performance with less or no help on a subsequent short-term transfer problem requiring the target concepts and skills. Reduction in mistakes in subsequent problems of the same type allows for evaluation of help effectiveness even though there is wide variability in the pattern of problems and performance across student users. These studies will allow us to collect data on how hard problems are for students, and on how effective the hints are for helping students solve problems.

A Bayesian network links the probability of answering a problem correctly, and the probability of mastering skills, following the philosophy of Mitrovic, Martin and Mayo (2002). We intend to have a probability of the student answering correctly each specific problem.

3 Social implications: Cognitive and gender differences

The previous project described a system to reduce gender differences in geometry problems. The next system specifically determined the effectiveness of different help for students of different genders and cognitive developments. Results showed that girls were more sensitive to the amounts of help than to the level of abstraction (e.g., the use of concrete objects to manipulate, (Figure 9), versus equations and procedures, Figure 10). Tests with 313 children showed that boys were affected by the level of abstraction, and were more prone to ignore help.

Past research suggests that good teaching will not be the same for students of different genders and cognitive developments. Flavell claimed that if a seven-year-old is in the stage of concrete operations², she or he should be given problems involving actual physical objects to manipulate rather than abstract symbolism or diagrams, thereby capitalizing on this student's strengths. There is evidence for the success of manipulatives such as base ten blocks in the early grades (Suydam and Higgins 1977; Fuson and Briars 1990). Concrete-operational students benefit from concrete materials (e.g. base-10 blocks) more than straight procedural numeric approaches to learn arithmetic. Other studies show that girls tended to use counting strategies, while boys tended to use memory retrieval and derived facts (Fennema, Carpenter, Jacobs, Franke and Levi, 1998). Studies of first-grade students found similar results (Carr and Jessup, 1997), claiming that the effect increased over the course of the school year. Thus, what constitutes effective math instruction may be somewhat different for girls and boys.

AnimalWatch is an Intelligent Tutoring System for basic arithmetic and fractions. It adapts its teaching to student characteristics such as gender and cognitive development. For example, concrete-operational students should benefit from concrete representations in the help for learning math instead of formal approaches, which may be more efficient for students of higher cognitive abilities. The results were that girls performed better in the following problem when the help was highly interactive, while boys improved more when the help had low levels of interactivity. Each hint in the system was labeled according to two levels of interactivity and abstraction, and all hint types were randomly given to students.

Mathematics problems. AnimalWatch tutors arithmetic with word problems about endangered species. Thus, it integrates mathematics, narrative and biology. Specifically, math problems are designed

to motivate students to use mathematics in the context of practical problem solving, embedded in an engaging narrative (Figures 11 and 12). Students are invited to work with scientists as they explore environmental issues endangering the animals. Students use mathematics to solve environmental and ecological problems as they observe, monitor and manage the endangered animals. AnimalWatch maintains a student model and makes inferences about the student’s knowledge as she or he solves problems. It increases the difficulty of the problems depending on the student’s progress, going from simple whole number addition problems to others that involve fractions with different denominators. Each problem given by the system will be both mathematically and pedagogically appropriate, based on the system’s estimation of the student’s knowledge, cognitive development and gender. When the student enters an incorrect answer, the system provides increasing levels of information.

AnimalWatch helps students learn fractions and whole numbers at a 4th-6th grade level. It provides mathematics instruction for each student based on a dynamically updated probabilistic student model. Problems are dynamically generated based on inferences about the student’s knowledge, progressing from simple one-digit whole-number addition problems to complex problems that involve fractions with different denominators.

**Student Model.** The student model continually updates an estimate of each student's ability and understanding of the mathematics domain. It generates problems of appropriate difficulty for each student and customizes help and instruction when students make mistakes. Students are given hints with little information first and richer explanations later if the former ones are not effective. The results of several evaluation studies with 10 and 11-year-old students indicate that AnimalWatch provides effective individualized math instruction. It has also been shown that AnimalWatch has a positive impact on students' own mathematics self concept and belief in the value of learning mathematics (Beck, Arroyo, Woolf and Beal, 1999).

When a student encounters a difficult problem, AnimalWatch initiates a tutoring interaction and provides hints classified along two dimensions, symbolism and interactivity. Hints progressively increase the amount of information given. The first hints provide little information, but if the student keeps entering wrong answers, AnimalWatch gives hints that will ultimately guide the student through the whole problem-solving process. We analyze how the number of mistakes the student has made changes from problem to problem after seeing a particular kind of hint. We looked for main and interaction effects for gender, cognitive development, hint interactivity and hint symbolism in predicting hint effectiveness.

The tutor took the student through a series of word problems about endangered species that it dynamically chooses from a large database of templates, which were instantiated with appropriate operands, depending on the student’s current proficiency level. AnimalWatch maintains a Bayesian overlay student model and makes inferences about the student’s knowledge as he solves problems. Based on these estimations about how students do in problems, AnimalWatch adjusted its problem selection so as to give the student problems of an appropriate difficulty. Student mistakes were corrected thanks to the help provided, so in the end AnimalWatch pushed the student forward, going from simple whole number addition problems to others that involve fractions with different denominators. AnimalWatch provided immediate help if the pupil entered a wrong answer.

The system first determined students’ cognitive development with a computer-based Piaget cognitive development test, which consists of ten multimedia-based tasks, which test for concrete and formal-operational thought. Students were classified into low and high cognitive development by splitting at the median score in this test. An instrument that measured self-confidence, math value and math liking...
was used. The number of mastered topics is a good estimation of how well students did with AnimalWatch, as it is highly dependent on the mistakes students made. The main predictors of the number of mastered topics were 1) the number of problems seen; 2) cognitive development; and 3) study (suggesting a study effect—suburban school children mastered more topics in general), but not the kind of help seen. A possibility for the help not being a significant factor in predicting the number of mastered topics is that different kinds of help may be best for students in different topics. For example, a high cognitive development student may be bothered by highly structured help in addition, while they may benefit from the same kind of help in fractions.

**Evaluation results.** The main predictors for errors made by the student were: 1) the total number of problems in the sequence; 2) the number of mistakes in the first problem of the sequence; and 3) an interaction effect between gender, help type and cognitive development ($F=6.5, p=0.000$). It was found that formal help was significantly worse than other kinds of help for low cognitive development boys. Also, formal help was significantly worse for low than for high cognitive development boys. Formal help produced significantly worse mistake change rates for boys of low cognitive development than for girls of the same cognitive development. Low cognitive development girls improved significantly less with reduced-help than girls of high cognitive development with reduced help. Also, girls of high cognitive development improve most with reduced help. AnimalWatch was effective in teaching arithmetic. The number of mistakes a student made on problems of a similar type continued to reduce, showing that they learned the topic (Figure 13).

Self-confidence and math liking increased significantly during the use of this system. On the other hand, there was a significant decrease in mathematics value ($p<0.046$), partly because of a ceiling effect. A significant interaction effect was found between gender and help-type ($F=3.15, p<0.046$). Girls’ math value was lowest after using reduced-help than any other kind of help. Also, girls using reduced help had lower math value than boys using reduced help. Reduced help produced an increase in math value for boys.

We found gender differences in the time that students spent on hints (independent samples t-test, $p<0.05$). On average, girls stayed 25% more time than boys within hints. This happened despite of the fact that, overall, girls mastered similar amount of topics than boys. The seconds spent per hint are very low. Part of the reason for this is that the first hints given to the student were short messages. However, spending less than 3 seconds in between responses doesn’t give enough time to give a second thought to the answer, suggesting students were intentionally ignoring help, probably searching for more informative hints. Aleven (2001) found similar behaviors. It is noticeable though that girls tend to display this behavior less (specially high cognitive development girls), while boys ignore help more (specially high cognitive development boys). We conclude that boys are being more selective about the help provided to them.

In sum, the behavior of girls and boys of same cognitive development was quite opposite. In general, the best help types for one gender are the worst for the other gender. Also, while girls’ math value and self-confidence is affected positively by the existence of intense help, boys’ math value is harmed. Some boys may have felt bothered by structured help, specially, when we consider they spent 25% less time at each hint than girls. Too much help when they didn’t need it would slow them down, probably they went through all the help while they could have figured it out by themselves. High cognitive development boys behaved opposite to high cognitive development girls: girls profited from
different amounts of help depending on the topic, while boys profited from intense help all over. Intense help also has the largest variety of help, as it contains all hints. We think they managed to decide what help they needed and search for it. We conclude the best type of help for low cognitive development boys would be to give the option all kinds of help available, as they seem to take it or leave it when necessary. We still believe that high levels of interactivity should be kept to a minimum for this group. We conclude that girls are more sensitive to the amounts of help than to the abstraction level. Thus, research efforts for girls should emphasize where support is needed (and provide it with plenty of interactivity when needed), as they don’t seem to ignore it. We conclude that research efforts on what help to provide low cognitive development boys should be on the representation used in the help, as they will tend to ignore help they don’t feel comfortable with.

4 Quality assurance. Improved learning and reduced costs

The previous systems have begun the process of customizing learning for the student, making learning engaging and exciting. The final system, though not as engaging, has been shown to reduce costs and improve learning. This is a Web-based learning system (OWL), which provides homework and interactive discovery modules for 35,000 students in 20 departments in 30 institutions across the country. Discovery OWL sets up learning environments where students discover basic chemistry concepts and are guided through exercises to deepen their understanding of the material (Figure 14). OWL Tutor provides intelligent tutoring on concepts well known to be difficult for most students.

Evaluation shows that use of OWL improves student learning, drives innovation in teaching, reduces cost for large enrollment and changes the way textbooks are used. Studies funded by two FIPSE grants and a National Science Foundation grant, have allowed us to demonstrate that use of OWL improves learning by keeping students on task with the instruction and allowing them to repeat assignments until they are confident of their abilities. Further, we found that OWL can drive innovation in teaching by allowing faculty to spend less time in class on material better taught through homework or intelligent tutors, and more time interacting with students on a deeper understanding of the material or introducing new material.

Student learning. OWL has a very positive impact on student performance. Our studies show that test scores improve significantly when classes adopt the Basic OWL system and also when students use the extensions to OWL such as intelligent tutors. Controlled pre-OWL and post-OWL studies in 2 introductory physics courses over 6 semesters have shown: 1) an increase in mean scores on midterm and final exams by an average of half a letter grade, 2) a decrease in the variance around the mean on these exams (overall performance is more concentrated around the increased mean, see Figure 15), and 3) greatest performance gains for the weakest students, those most likely to fail or drop the class.\(^3\)

\(^3\) Exams were controlled so that they were very nearly equivalent across the three years. Exams were collected after their administration so that students couldn’t share them from section to section or year to year. Analysis of SAT Math scores shows that student populations were equivalent in ability from year to year.
The group of students who completed the OWL homework scored 18% higher on exams on average than those who did not, and none of the «high» group scored below 50%. It comes as no surprise to most educators that students who do their homework perform better on exams.

**Teaching innovation.**

Evidence from surveys, focus groups, and observed changes in teaching practice show that use of OWL is driving innovation in the way instructors teach. For example at UMass we have completely eliminated weekly quiz/discussion sections and have replaced them with OWL and the Chemistry Resource Center (a tutoring/computer laboratory, staffed by chemistry teaching assistants and faculty). Two-thirds of the faculty in the Chemistry Department were involved with teaching and grading discussion sections, hence we have seen a significant savings in faculty time, with no decrease in the quality of instruction, by adopting this model. Faculty who have taught with OWL describe its main benefits as:

- OWL improves class preparedness and allows lecturers to deal with a heterogeneous population in terms of providing a more common level of preparedness.
- OWL allows faculty to reduce in-class drill and review of homework problems and spend more time on concepts.
- OWL alleviates the tremendous burden of managing regular homework assignments in courses and sections with hundreds of students.
- OWL has changed the way basic material is handled in honors classes.

**Reduce costs.** We have determined that OWL requires fewer instructional resources while at the same time improving student achievement, i.e. *better instruction at lower cost*. From analyses of instructor time, support staff, and facilities use in the Physics Department, we can conclude that more than 100 000$/year has been saved by adopting OWL in that department, and that number is still growing as more physics classes adopt OWL.

**Use of Textbooks.** In general chemistry classes at UMass instructors assign both OWL and textbook homework problems. This is because the text has questions not covered in current OWL modules, and OWL has questions not covered in the text. On exams, students frequently perform better on questions drilled in OWL than on those only covered in the text. This is partly because they either don't complete the text questions, which are not graded, or they do not work enough similar questions to thoroughly learn the material. When asked about this, students report that they find the feedback in OWL more helpful than the text when learning how to do homework, and the ability to rework similar, but not identical problems in OWL, gives them the chance to be sure that they understand (students frequently request that all of the homework be assigned in OWL). We also find that as the number of discovery modules and tutors in OWL grows, students and instructors alike depend increasingly on OWL and less on the text. Faculty report that they sometimes assign OWL homework in areas they don’t cover in class with more assurance that students will succeed than if they just assigned a chapter from the text.

This study compared the effect of OWL homework (Web-based homework) and paper-and-pencil homework (paper-based homework) on student achievement as measured by exam performance. Various offerings of two large introductory physics sequences were examined over a three-year period, with some courses taught with paper-based homework and some with Web-based homework. We found that Web-based homework offerings led to higher overall exam performance; the mean difference between Web-
based homework and paper-based homework courses was about one-third of a typical exam standard deviation. In one comparison, where there were matched final exams, the difference in exam performance was statistically significant. In addition, significant cost savings were realized in moving from paper-based homework to Web-based homework. Overall, we found that replacing paper-based homework with Web-based homework led to higher physics achievement as measured by exam performance. The difference in performance was about a third of a typical exam standard deviation for a given class. Our findings indicate that students who perform well on homework generally attain higher achievement scores. The performance gaps in exam scores between Web-based homework and paper-based homework were similar for low and high SAT-math groups, for low and high homework score groups, and for low and high exam performance groups.

5 Conclusions

Educational institutions are caught in a rush to put courses «online» (or at least appear to have done so) though many have not first questioned the possibilities afforded by the medium, nor sought to understand the human issues in the process. The goal of interactive media in education should not be to force learners into different molds, but rather to remove the mold altogether. It is a chance to liberate individuals through exploration and construction that produce learning outcomes.

Additionally, we are moving to a state in which working and learning are increasingly the same activity. The demand for learning as part of a lifelong challenge creates a demand in the workforce for education that will exceed the capability of traditional institutions. This creates an opportunity for new intermediaries and learning agents that are not part of the traditional, formal education system. Such opportunities are likely to be enhanced by computer technology.

We also have an unprecedented opportunity to use computer teaching to research learning and teaching. We can study the effect of a specific teaching strategy with a classroom of students, by regulating the teaching in a system, e.g. problem/hint selection, carefully tracking student performance and evaluating performance crossed by gender or cognitive development. We can also study individual learning strategies by tracking individual errors made crossed by student background or cognitive development.

The systems described in this paper push the frontier of electronic teaching toward more responsive and engaging tutors. They introduce new pedagogy, address inequities in the classroom and show improved learning with reduced costs. They often develop a model of the user, customize their feedback to the student, and use multimedia to engage students and demonstrate visual problem solving.

We have built systems to help all students perform to the best of their abilities. «Wayang Outpost», an Intelligent Tutoring System, prepares high school students for passing the high stakes achievement tests used for college entrance. It uses multimedia and tutoring decisions to effectively tutor every student and address factors that make females score badly in these tests.

We explored the effectiveness of different help for grade school students of different genders and cognitive developments, and concluded that girls were more sensitive to the amounts of help fitting their needs than to the level of abstraction. On the other hand, boys were affected by the level of abstraction, and were more prone to ignoring help.

Our Web-based homework system is used in 20 college departments including physics, chemistry and geology. This system has been adopted by nearly 35,000 students/semester and 40 colleges. Research provides solid evidence that electronic homework results in reduced costs, improved student grades (sometimes adding a full letter grade to student scores), innovation in teaching, and changes in the way textbooks are used. OWL improves learning by keeping students on task with the instruction and allowing them to repeat assignments until they are confident of their abilities.

Computer-based systems can also teach specific learning skills. We described an inquiry learning or scientific reasoning system that tracks, analyzes and then comments on a student’s selection of data and creation of hypotheses and inferences. We expect to identify the strategies students use to generate hypotheses and explore data.
6 Bibliography


Acknowledgements

Research described in the first section was funded in part by the National Science Foundation under grant no. DUE-0127183, « Inquiry Tools for Case-based Courses in Human Biology », M. Bruno, PI. And partly by the U.S. Department of Education, « Expanding a General Model of Inquiry Learning », Fund for the Improvement of Post Secondary Education, Comprehensive Program, #P116B010483, B. Woolf, P.I.

Wayang Outpost software, Section 2, is being developed with support from the National Science Foundation’s Program for Gender Equity, grant no. HRD-012089 to C. Beal. AnimalWatch research, Section 3, was also supported in part by grants from NSF’s Program for Gender Equity, HRD-9555737 and HRD-9714757 to C. Beal.

Research regarding the OWL system, described in Section 4, was funded in part by the U.S Department of Education, FIPSE Comprehensive Program grant P116B70834 to W. Vining, « A Web-based Tutor for General Chemistry », and FIPSE grant P116R980038 « Controlling Costs of Higher Education », to B. Woolf; and by the National Science Foundation Course and Curriculum Development grant DUE-9653064, R. Day, PI. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.