DOCUMENT RESUME

ED 409 217 SE 060 483

| | |
|---|---|
| AUTHOR | Steuck, Kurt; Miller, Todd M. |
| TITLE | Evaluation of and Authentic Learning Environment for Teaching Scientific Inquiry Skills. |
| PUB DATE | Mar 97 |
| NOTE | 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997). |
| PUB TYPE | Reports - Evaluative (142) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Computer Assisted Instruction; *Computer Software; Computer Uses in Education; Educational Technology; Evaluation; *Inquiry; *Science Experiments; Scientific Methodology; *Scientific Research; Secondary Education; Student Research |

ABSTRACT

Instruction in Scientific Inquiry Skills (ISIS) is part of a seven-year Air Force effort--the Fundamental Skills Training Project--to design, build, evaluate, and transition advanced computer-aided instruction to the education community. This report describes ISIS 2.0 and presents the results of an initial field evaluation of the software during the 1995-1996 academic year. The description of cognitive apprenticeship as implemented by ISIS includes modeling expert knowledge and performance skills, coaching as a collaborative effort, structuring, fading authentic knowledge and skill performance, and reflection and articulation. Results are classified according to overall test scores, development of a research question, generation of a hypothesis, design of an experiment, conduction of an experiment, drawing a conclusion, acceptance or rejection of hypotheses, and domain knowledge. Findings indicate that ISIS improves students' scientific inquiry skills more than traditional large-class instruction. Students' skills and domain knowledge grew in a linear fashion as experience with the tutoring system increased. This pre- to post-test gain was found to be larger for inquiry skills than for the domain knowledge. Contains 16 references. (JRH)

# Evaluation of an Authentic Learning Environment for Teaching Scientific Inquiry Skills

**Kurt Steuck**

**Armstrong Laboratory**
**Human Resources Directorate**
**7909 Lindbergh Drive**
**Brooks Air Force Base, TX 78235-5352**

**Todd M. Miller**

**Mei Technology Corp.**
**8930 Fourwinds Drive #450**
**San Antonio, TX 78239**

**Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL March 1997**

# PREFACE

ISIS (Instruction in Scientific Inquiry Skills) is part of a 7-year Air Force effort -- the Fundamental Skills Training project -- to design, build, evaluate, and transition advanced computer-aided instruction to the education community. This article does not necessarily reflect the opinions or policies of the U.S. Air Force or any other government agency.

This report describes ISIS 2.0 and presents the results of an initial field evaluation of the software during the 1995-1996 academic year. Many people contributed to the development of ISIS. The authors express their gratitude to the high school teachers who served as subject matter experts (Dr Carolyn Pesthy, Dr George Williams, Patricia Jackson, Dave Bordelon, and Steve Holbrook); the programmers who developed the software (Keith Brown, James Johnson, Marcia Cromley, and Doug Estrumse); and the research assistants who collected, tabulated, and analyzed the data (Nick Meyer and Jenifer Wheeler). We especially acknowledge the generous sharing of time and talent and continued support of the following individuals: Dr. Wes Regian, Senior Scientist for the Intelligent Training Branch (AL/HRTI) and Lt Col Jim Parlett.

Funding for the FST project occurs through partnerships with several organizations. Our current partnerships include a Cooperative Research and Development Agreement between the Human Resources Directorate of Armstrong Laboratory and The University of Texas at San Antonio. In addition, Armstrong Laboratory has joined with the three other Air Force "super-laboratories" -- Rome Laboratory (Rome, NY), Wright Laboratory (Dayton, OH), and Phillips Laboratory (Albuquerque, NM) -- in a Memorandum of Understanding to support this research and the public school test facilities used by researchers in the project.

## TABLE OF CONTENTS

## LIST OF FIGURES

# Evaluation of an Authentic Learning Environment for Teaching Scientific Inquiry Skills

## 1.0 Introduction

Recent progress in both theory and practice suggests that an authentic, constructivistic approach to science education is central to the development of meaningful conceptual models of scientific phenomena (National Science Education Standards, 1996; Glynn & Duit, 1995; Rutherford & Ahlgren, 1990; Yager, 1995). This finding, coupled with research demonstrating the effectiveness of computer-based instruction (CBI) across student populations and knowledge domains (e.g., Kulik & Kulik, 1991; Kulik & Kulik, 1985; Kulik, Bangert, & Williams, 1981) highlights the need to test authentic, adaptive systems for science education.

This study, conducted under the umbrella of the USAF's Fundamental Skills Training (FST) project[1], seeks to evaluate the instructional efficacy of the *Instruction in Scientific Inquiry Skills* tutor (ISIS), an intelligent tutoring system (ITS) for teaching scientific inquiry skills in the context of ecology and biology. This document begins by briefly describing the software and the philosophy underlying its design. The crux of the paper, however, outlines the study methodology, research findings, and discusses implications for future ITS design.

### 1.1 Description of ISIS

ISIS is a simulation-based ITS designed to teach junior and senior high school students scientific inquiry skills and substantive knowledge in ecology. The first goal-- to teach students the skills underlying scientific inquiry-- requires that the learner engage in a series of problem-solving exercises whose successful completion requires the student to: 1) formulate an important and testable research question, 2) generate a relevant hypothesis, 3) design an experiment to test that hypothesis, 4) conduct the experiment in a simulated ecosystem, 5) draw an appropriate conclusion based on experimental data, and 6) accept or reject the hypothesis accordingly. The second goal-- to teach students ecological concepts and their interrelationships-- is accomplished by framing the problem-solving assignments in real-world domains of ecology by requiring students to perform these exercises in a simulated ecosystem. ISIS teaches ecology concepts in areas including biomes, abiotic factors of plant growth, biotic factors in ecosystems, human activities, and ecology principles.

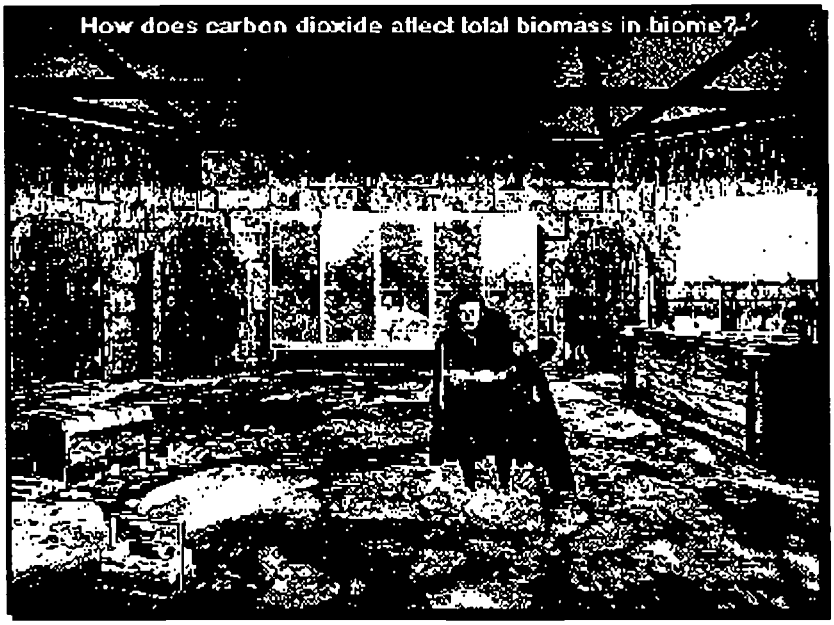
ISIS is set in a gaming context in which students are to "buy back the biomes" from the Grim Reaper. An introductory tour shows major biomes to students and challenges them to save the planet. The Grim Reaper shows how Earth is being damaged. A friendly wizard tells students to save the planet by applying knowledge and skills. Upon completing this tour, the students learn scientific inquiry skills through Skill Instructional Modules (SIMs). Each of the inquiry skills listed above is covered in one SIM. Students learn about ecology through Domain Instructional Modules (DIMs) covering major biomes, abiotic factors of plant growth, biotic factors in an

---

[1] FST, funded by the Air Force Armstrong Laboratory's Intelligent Training branch, is a multi-year effort to develop, evaluate, and transfer adaptive training technologies to public education and, where appropriate, to industry under federal technology transfer guidelines.

ecosystem, and the effects of human activities in ecology. After completing a SIM, students apply their newly learned knowledge and skills in a simulated medieval castle (See Figure 1). The castle includes Igor, a lab assistant, and a wizard who gives advice. Students use a set of "magic mirrors" to select a research question. Along one wall is an equipment room which contains equipment students use to measure their independent and dependent variables. There is also a library which students can access the skill instruction and domain readings, a glossary, and interface help. Students use lab equipment (e.g., terrarium) on a desk to perform the scientific skills (e.g., design an experiment to test their hypothesis). To complete the gaming aspects of ISIS, students are given points for completing the skill instruction, domain readings, and research activities. They use these points to "buy back biomes" from the Grim Reaper.

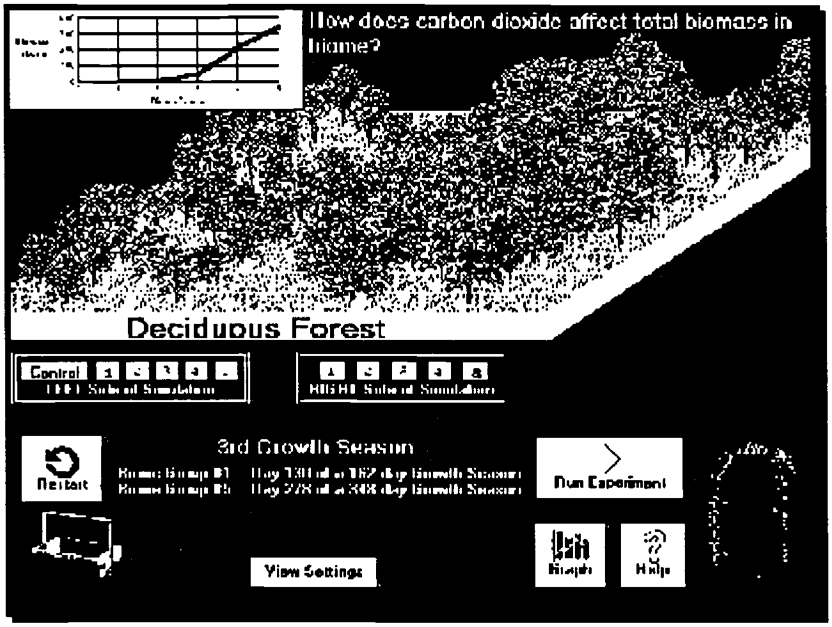
**Figure 1. The Castle**



How does carbon dioxide affect total biomass in biome?

Once they have completed the first skill instruction on generating research questions and hypotheses, students are transported to the castle where they begin to perform scientific inquiry skills in an adaptive and supportive environment. Students are slowly guided through the scientific process in a step-wise fashion. Students first select a research question from over 300 available questions presented in simplified concept maps. They then generate a hypothesis by selecting an independent variable (IV) and dependent variable (DV) that corresponds to the research question they have chosen. After having chosen the IV and DV, they pick a graphical representation of the relationship between the two variables. In designing the experiment, students specify the IV and DV, the units used to quantify those concepts, and the equipment needed to measure those units. At the core of ISIS is a set of simulations (see Figure 2) of ecological niches or biomes (e.g., coniferous forest, desert). Here, students conduct their research by manipulating abiotic variables in accordance with the research question, hypothesis, and design specified earlier. Once the simulation is set into motion, students observe the impact of these manipulations on various graphical indices. After running the simulation, students are required to draw conclusions based on the results of the simulation and then accept or reject their initial hypothesis.

**Figure 2. Sample Simulation**



How does carbon dioxide affect total biomass in biome?

1.2 Constructivism in Science Education

Curricular emphasis, particularly in the realm of science education, has slowly moved from an *objective* model of instruction to a more *constructive* perspective (e.g., Carey & Smith, 1993). Objectivism posits that the goal of instruction is, in essence, "to map an external reality onto learners" (Jonassen, 1991). As a result, learners often view "knowledge as arising unproblematically (and directly) from sensory experiences and see knowledge as simply the collection of many true beliefs" (Carey & Smith, 1993). Instruction, then, is nothing more than the act of analyzing a given content domain and then conveying that information to learners.

Constructivism, on the other hand, posits that instruction should focus on activities that facilitate knowledge construction by individual learners. By providing appropriate, authentic learning situations, instructional systems (teachers and computers) are better able to help learners construct meaningful and conceptually valid representations of the external world (Jonassen, 1991, Jonassen, 1991a). Under this paradigm, instructors and instructional systems provide the learner with the opportunity to interact with context-rich learning environments while helping to manage cognitive load by gradually increasing the complexity of the conceptual models employed for instruction. In turn, students are able to construct mental models which more closely parallel the tutor's implicit conceptual model. As a result, learners have a network of knowledge which is more readily activated in new learning and performance situations.

This perspective is especially germane to science education. The traditional approach has emphasized the teaching of *facts* over process skills. When process skills are taught, they are often taught under highly artificial, contrived conditions. Students, for example, are often required to conduct lab experiments, but standard curricular approaches (e.g., textbooks) usually pre-specify the *correct* hypothesis, methodological approach, and results so that all the learner is required to do is follow a cookbook approach to scientific inquiry (Schauble, Glaser, Duschl, Schulze, & John, 1995). In practice, however, intuitions, hypotheses, and theory guide the process. Often times, several methodologies are available to test a given hypothesis and the same set of results can be interpreted in different ways by different researchers. In other words, science does not proceed in a lock-step manner; rather, scientific inquiry is a dynamic process and scientific phenomena are multi-faceted entities which can be viewed from divergent perspectives.

### 1.3 Cognitive Apprenticeship

Cognitive apprenticeship is an instructional approach which embraces constructivism, but does not inherently prohibit the simple transmission of information from instructor to student (Collins, Brown, and Newman, 1989; Collins, Hawkins, and Carver, 1991; Hsieh, Miller, Hicks, & Lorenz, 1993). Rather, cognitive apprenticeship is a collaborative approach to instruction in which tutor and pupil interact to facilitate knowledge construction in the ways which are most relevant to the learner. The key tenet, however, has to do with the notion that *knowledge is best constructed from experience.* As a result, instruction based on pure transmission or drill-and-practice (or discovery) is not ideal (each has limitations). Instead, instructors should create situations in which learners are able to derive insight on their own or under limited guidance.

This approach is especially relevant to computer media because a thoughtfully designed system can provide contextually rich, authentic learning environments for a wide range of students over a wide range of disciplines (Hsieh, Miller, Hicks, & Lorenz, 1993). The following section describes how ISIS implements the principles of cognitive apprenticeship.

### 1.4 Cognitive Apprenticeship as Implemented by ISIS

The key features of cognitive apprenticeship are: modeling (showing), coaching (telling), authentic performance of knowledge and skills (doing), student reflection and/or articulation (metacognition). The next few sections will discuss how these features were implemented in ISIS.

#### 1.4.1 Modeling Expert Knowledge and Performance Skills

Modeling, and apprenticeship in general, refers to the act of an expert performing or demonstrating knowledge and skill for a novice. The idea is that novices, regardless of domain, learn best initially by observing the performance and/or thought processes of experts (Collins, Brown, and Newman, 1989). In ISIS, modeling is accomplished by first requiring users to work through a series of interactive CBI lessons designed to teach them the content and skills necessary to perform specific scientific inquiry skills. In these exercises, students are presented with information, shown expert solutions (e.g., what is a testable hypothesis?, what graph is best suited for a particular type of data?). Then, students are required to demonstrate similar skills and competencies by manipulating the interface to perform a wide-range of tasks (e.g., generating hypotheses, selecting testable hypotheses). If they do not exhibit mastery of the concepts in question, they are directed to a remedial loop in which they receive additional coaching.

### 1.4.2 Coaching as a Collaborative Effort

Coaching, by definition, refers to a collaborative activity in which teacher and pupil interact to reach a common objective: skill or knowledge mastery. The teacher requires the student to perform certain skills or demonstrate content mastery in authentic tasks. If the student does not display mastery, then the teacher resorts to describing and modeling the appropriate responses.

ISIS performs these functions throughout the system, but primarily in the instructional modules and the adaptive help system. In the instructional modules, students who reach a threshold of errors are given directive feedback to remediate misconceptions. When they reach the adaptive environment (i.e., the castle) and begin skill performance, errors are remediated by an assignment-independent mechanism which identifies the source and direction of their error. This mechanism progressively gives more detailed help. The first two help statements indicate the location of the error (e.g., advice statement: your design is not consistent with your hypothesis; your independent variable is not the one you specified in your hypothesis) The third statement gives a way to remediate the error (e.g., Compare the your design with your hypothesis). The final help statement gives the solution to the impasse (e.g., Your independent variable is rainfall) so that students are not stuck at any one point for a lengthy period of time. End-of-year interviews with teachers revealed that the teachers thought students were not "abusing the wizard" by asking for help too often.

### 1.4.3 Structuring

Another component of cognitive apprenticeship is structuring of lesson content and subsequent fading of the structure (Collins, Brown, and Newman, 1989). ISIS structures students' interaction with ISIS by interleaving modeling of skills and eliciting students' performance of those skills in the intelligent environment. For example, students complete the instructional module on generating hypothesis before completing assignments in which the skill is required. Igor, the lab assistant structures the learning session by guiding the students through the inquiry skills one at a time in sequence. Igor also gives feedback as to the appropriateness of the students' actions after the students complete each skill.

### 1.4.4 Fading

Fading is accomplished in ISIS via the wizard's instructional statements by tailoring the advice to the student's needs and level of proficiency. This coaching mechanism is sophisticated in that it differentiates between students of different proficiency levels (based on task performance) and tailors the coaching accordingly. Students of lower proficiency are given more directive statements earlier in the process while higher proficiency students receive leading questions.

### 1.4.5 Authentic Knowledge and Skill Performance

The cornerstone of Constructivism and cognitive apprenticeship lies in the instructor's or instructional system's ability to create realistic environments in which student can practice his or her skills (again modeling and coaching are options available here) AND the ability of the system to respond appropriately to student errors. Authentic environments/situations are those which mimic the real-world with enough fidelity to bypass the students inherent tendency to compartmentalize the world into academic vs. real-world experiences. The unfortunate side-effect of traditional classroom instruction is that learners often are not able to make connections between the two (can solve algebra word problem, but can't solve checking account problem with one unknown, e.g., a missing check). By situating instruction and practice in viable scenarios, students are often better able to transfer potentially inert knowledge to new situations.

The adaptive working environments is a real strength of ISIS. Based on simulations and carefully designed interfaces, ISIS is able to provide the student a micro-world for scientific inquiry. Cognitive load is minimized by not presenting students with inquiry skills with which they have not yet received tutoring on (or demonstrated mastery of), providing hints for partial solutions, or giving answers to and suggestions for remediation to those bogged down by the process.

### 1.4.6 Reflection and Articulation

Articulation, as an instructional tool (strategy), refers to the idea of creating an environment in which students rate or judge their work with regard to a standard (Collins, Brown, and Newman, 1989; Collins, Hawkins, and Carver, 1991). In ISIS students reflect back either on their own performance or some aspect of the domain content after completing an assignment. This helps students develop a stronger understanding of the ecology concepts under study and their own abilities in performing the scientific inquiry skills. For example, after working through an exercise dealing with the effects of $CO^2$ on leaf biomass, the user may see a question in his or her notebook on (e.g., What steps did you use when conducting your experiment?). These questions are designed to have students apply their newly learned knowledge further or answer questions about their own performance of scientific inquiry. These types of activities help to cement emerging knowledge by forcing the learner to think about the steps they used to arrive at a particular solution.

## 2.0 Methods

### 2.1 Subjects

Seventh, ninth, and tenth grade students (n = 1547) enrolled in 84 sections of introductory biology or life science at fifteen junior and senior high schools in five states across the nation were the subjects in this study. The sample of students across the schools was demographically diverse. The students participated in this research as part a part of their normal instruction in biology or life science classes.

Prior to analysis, a uniform set of cleaning rules were applied to the experimental data. Subjects had to meet three criteria in order to be retained. First, they had to have matching pre- and posttest data. Second, treatment subjects had to have on-line user data (e.g., errors made, steps taken) available for analysis. This was to standardize the research sample across all data analyses including some that are not reported in this paper. Finally, all subjects were required to score above chance for inclusion in the final sample. This was done because test proctors noticed that a small number of subjects either refused to take the test or did so in a haphazard (i.e., random) fashion. After these procedures were applied to the data, the sample size was reduced to 1547 subjects.

## 2.2 Design

There were two research questions addressed during the 1995-1996 academic year study. Both deal with the instructional efficacy of ISIS relative to traditionally used methods (i.e., the traditional classroom). One focuses on learning inquiry skills, the primary goal of ISIS. The second focused on learning the substantive or declarative knowledge taught in ecology classes. Specifically, the research questions were: 1) does ISIS help to improve a students' ability to perform scientific inquiry skills over and above traditional methods of instruction? and 2) does ISIS help to improve students' knowledge of ecology concepts over and above traditional methods of instruction?

To address these research questions, a quasi-experimental, two-group, pretest-posttest design was used (Campbell and Stanley, 1963). A quasi-experimental design was selected because the research team could not randomly assign teachers to treatment or control conditions nor could we assign students to sections of biology.

## 2.3 Instruments and Procedures

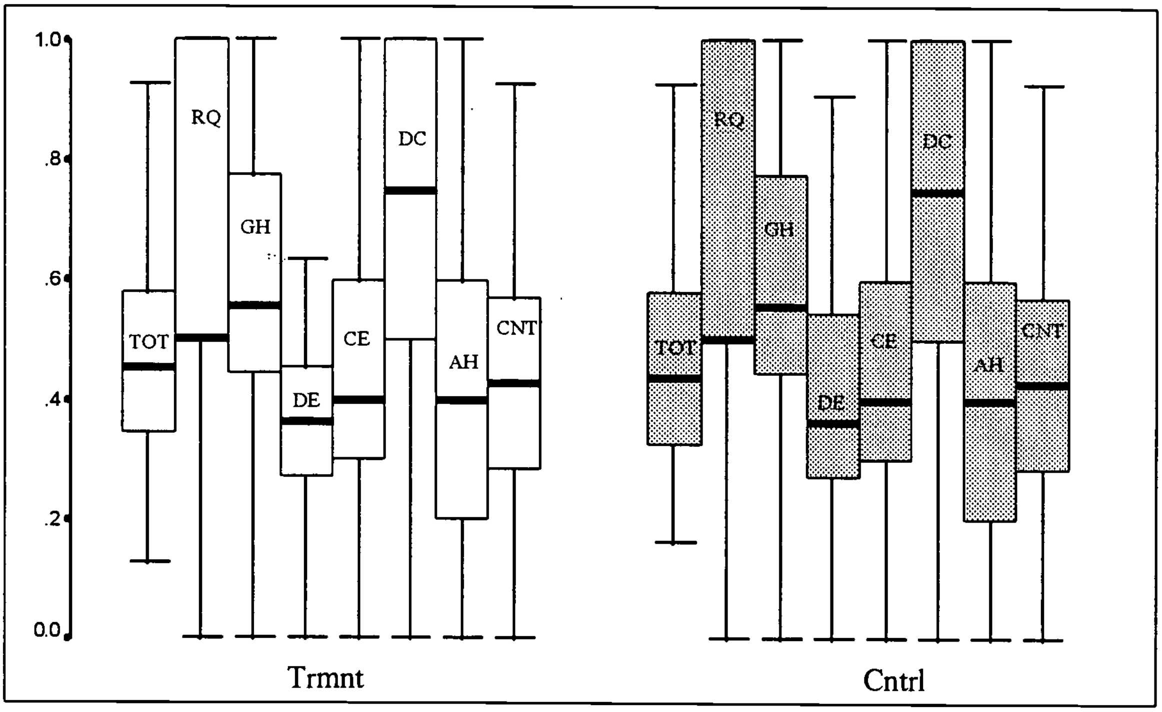
Two parallel forms of a psychometrically-sound (coefficient $\alpha$ was .92), 55-item pretest covering inquiry skills and knowledge of ecology concepts was developed for pre- and posttesting. Each form consisted of subscales covering domain knowledge and each of the scientific inquiry skills (e.g., generate a hypothesis). There were two questions assessing students' knowledge of research questions whereas four to 11 questions were used to assess each of the other skills. In addition, a short survey assessing attitudes towards school and learning was developed for this research.

The pretest and survey were given to all students during the first six weeks of the academic year. At the time of pretesting students were not aware of the specific purpose of the study or to which group they were assigned. After taking the pretest, the experimental students attended a computer lab approximately one day every two weeks during their normal class time for a total of about 18 hours over the course of the school year. Control group subjects went to their regular biology or life science class. Non-treatment control teachers were expected to cover the scientific method and ecology as part of their normal curriculum while treatment teachers were explicitly instructed to allocate as little time as possible to these topics. At the end of the academic year, students were given a posttest measure parallel to their pretest version.

# 3.0 Results

Because a quasi-experimental design was used, the authors could not ensure/assume equivalency between the treatment and control groups at pretest. Therefore, prestest score distributions were examined in order to ascertain the nature of these differences. Figure 3 presents a series of paired boxplots for each of the 8 scores derived from the skills and domain knowledge test. The plots on the left side of the figure represent ISIS subject scores while the corresponding plots on the right side of the graph represent control group scores.

**Figure 3. Boxplots for Pretest Subscale Scores by Condition**



*Legend. TOT is total score. RQ is research question subscale score. GH is generate hypothesis subscale score. DE is design experiment subscale score. CE is conduct experiment subscale score. DC is draw conclusion subscale score. AH is accept or reject hypothesis subscale score. CNT is domain knowledge subscale score.*

These paired boxplots indicate similar pretest score distributions across all of the scale scores. As can be observed, when looking across conditions, all of the scores have wide intervals, similar medians. Looking across the scale scores themselves, distinct differences appear however.

A MANOVA was used to test for differences between conditions on each of these 8 scores. MANOVA was use to control for correlations between dependent variables, which ranged from .32-.79. This analysis revealed only 1 significant univariate difference among pretest scores between conditions. This difference occurred for the *draw conclusion* subscale ($p = .022$) in favor of the treatment group (roughly a 3% difference). Generally, these boxplots suggest that treatment and control subjects were equivalent and therefore derived from a uniform population. Nonetheless, the researchers felt it better form to control for the slight pretest differences in aptitude. Therefore, ANCOVA with simple contrasts were used to test for main effects of instructional efficacy.

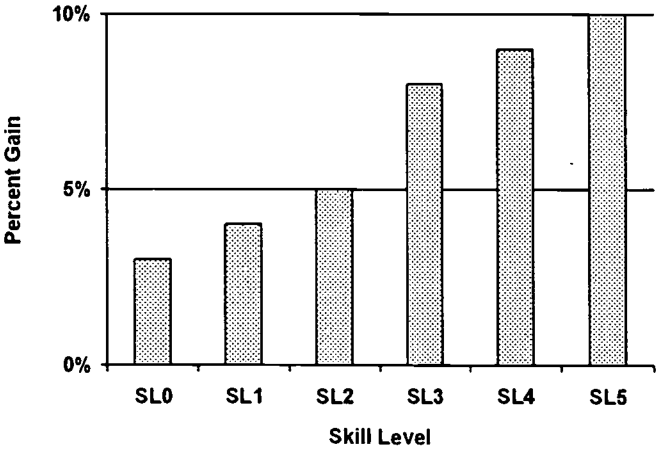
### 3.1 Instructional Efficacy

In order evaluate the instructional efficacy of ISIS, the treatment groups were partitioned into 5 groups corresponding to the *skill level* (SL) attained by each student. Since control group students did not use the tutor, they were assigned to the group *SL0* ( n = 788). The other groups were: *SL1* (ISIS students who received instruction and guided practice on generating hypotheses, n = 61), *SL2* (ISIS students who received instruction and guided practice on previous skill plus designing experiments, n = 136), SL3 (ISIS students who received instruction and guided practice on previous skills plus conducting experiments, n = 176) , *SL4* (ISIS students who received instruction and guided practice on previous skills plus drawing conclusions and accepting or rejecting hypotheses, n = 166), and *SL5* (ISIS students who received instruction plus additional practice on all skills, n = 226). Thus all experimental subjects were classified according to the level of instruction and experience that they had with the scientific inquiry skills as defined by ISIS.

### 3.1.1 Overall Test Scores

Overall test score refers to the percent correct out of 55 items covering scientific inquiry skills and substantive knowledge in ecology. Pre- to posttest scores gains by *SL*, shown in Figure 4, were as follows: *SL0* = 3%, *SL1* = 4%, *SL2* = 5%, *SL3* = 8%, *SL4* = 9%, *and SL5* = 10%. These gains correspond to pre- to post-test effect sizes ( $ES = \frac{Mean_{pst} - Mean_{pre}}{SD_{pre}}$ ) of .18, .29, .33, .53, .64, and .71 standard deviations respectively.

**Figure 4. Overall Score Gains by SL**



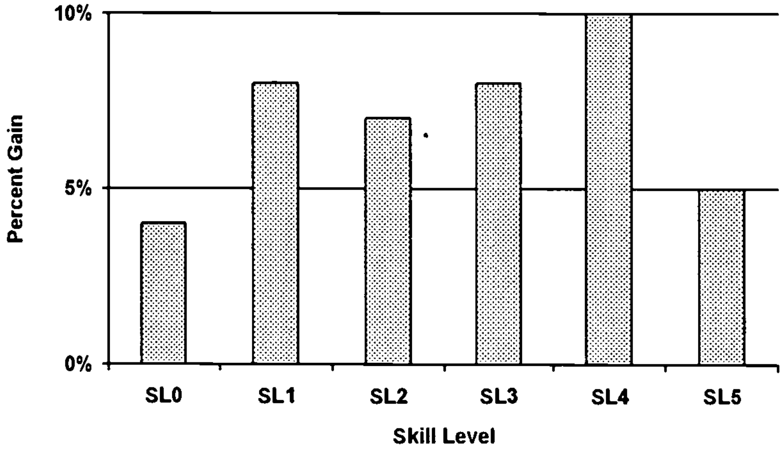
As a result of unequal pretest means between *SL* categories, ANCOVA was used to statistically control for these differences. The ANCOVA, using overall pretest score as a covariate and overall posttest score as the dependent variable, was significant F[(5, 1540) = 22.89, p < .001]. Overall, the model of interest was able to account for about 59% of the score variance. Prior knowledge accounted for about 52% of this variance while the SL factor added the remaining 7%.

As a follow-up, simple contrasts-- using the control (*SL0*) as the reference group-- indicated that students needed to reach *SL3* and beyond in order to significantly outperform the control. Effect sizes between SL1 through SL5 and SL0 were adjusted for the impact of prior knowledge ($ES = \frac{AdjSL0Mean_{pst} - AdjSL1-5Mean_{pst}}{SD_{SL0}}$, where $\text{AdjSL0Mean}_{pst}$ was the adjusted control group posttest mean, $\text{AdjSL1-5Mean}_{pst}$ the adjusted posttest mean for each of the 5 treatment groups, and $SD_{SL0}$ the control group standard deviation). These ranged from -.06 (SL1) to .47 (SL5) indicating that subjects with exposure to all skill levels scored about one-half a standard deviation above their control counterparts. Put another way, the average SL5 student outscored about 67% of all control subjects.

### 3.1.2 Developing a Research Question

As mentioned previously, the overall pre-posttest gain score was partitioned into a set of seven different subscales (e.g., developing a research question, generating a hypothesis). These indices were mutually exclusive of one another. In other words, each subscale was a composite of different test items and in forming these composites, all test items were used. The *research question* (RQ) composite was formed by summing a number of questions generally asking students to determine suitable research questions based on a scenario. Figure 5 shows pre- to post-test RQ subscale score gains by SL group. The within-skill level effect sizes were .11, .24, .18, .38, .31, and .16 respectively.

**Figure 5. RQ Subscale Score Gains by SL**



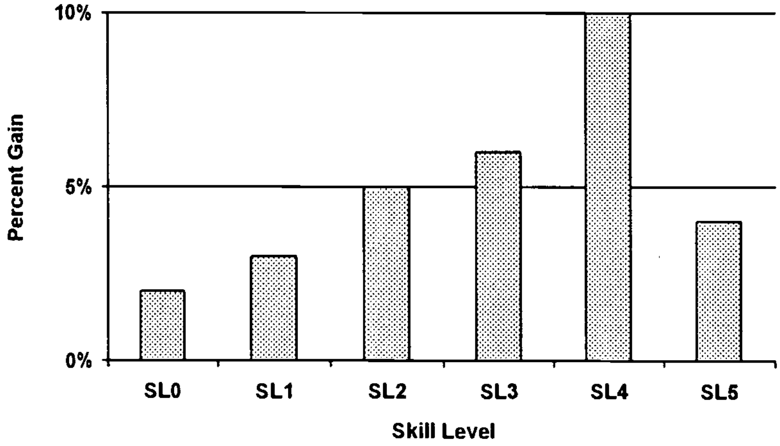
The ANCOVA results indicate a significant main effect of skill level F[(5, 1540) = 5.60, p < .001]. This SL effect was much smaller ($\eta^2 = .018$) than for the overall score, as was the impact of the covariate ($\eta^2 = .086$). This may be in part attributable to the relatively high RQ pre-test scores (see boxplots, Figure 3). As a result, the overall model only accounted for about 11% of the observed variance. Prior knowledge described about 9% of this variance while SL made up the remaining 2%.

Again, simple contrasts were used to examine mean differences between SL0 and SL1 through SL5. As indicated in Figure 5, treatment subjects needed to reach SL3 to achieve results significantly better than control at $p < .05$. Adjusted effect size measures were *SL1* vs *SL0* = -.10, *SL2* vs *SL0* = -.02, *SL3* vs *SL0* = .17, *SL4* vs *SL0* = .32, and *SL5* vs *SL0* = .21.

### 3.1.3 Generating a Hypothesis

The *generating hypothesis* (GH) composite was derived from a set of questions which required the students to select a important and testable hypothesis in response to a scenario. Figure 6 shows score gains by SL. Control group gained about 2% while the SL1-5 groups gained from 3% to 10%. Within-group effect size measures were .09, .13, .27, .29, .38, and .21 respectively.

**Figure 6. GH Subscale Score Gains by SL**

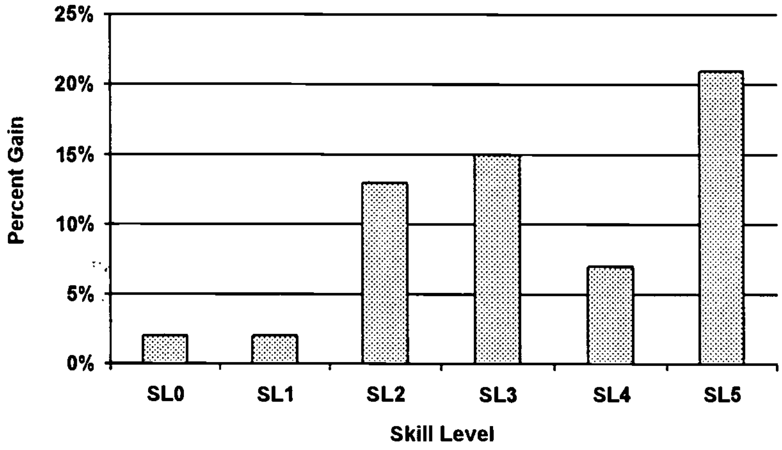


ANCOVA results indicate significant main effect of *SL* F[(5, 1540) = 7.70, $p < .001$]. The covariate effect F[(5, 1540) = 495.87, $p < .001$]. was significant as well. The total model accounted for roughly 28% of the score variance-- with 24% attributable to prior knowledge and 4% attributable to the SL factor. Simple contrasts revealed significant differences between SL0 and SL3, SL4, and SL5. Treatment subjects needed to reach SL3 to achieve results significantly better than control at $p < .05$. Adjusted between groups effect sizes were -.09, .07, .21, .35, .27 for SL1-5 when compared to SL0.

### 3.1.4 Designing an Experiment

The *design experiment* (DE) composite was based on a set of questions that dealt with subskills like identifying independent and dependent variables, and selecting appropriate methods (e.g., sampling) for designing quality experiments. Pre- to post-test gains were 2%, 2%, 13%, 15%, 7%, 21%. Pre- to post-test effect sizes were .10, .11, .77, .88, 1.00, and 1.11 respectively.

**Figure 7. DE Subscale Score Gains by SL**



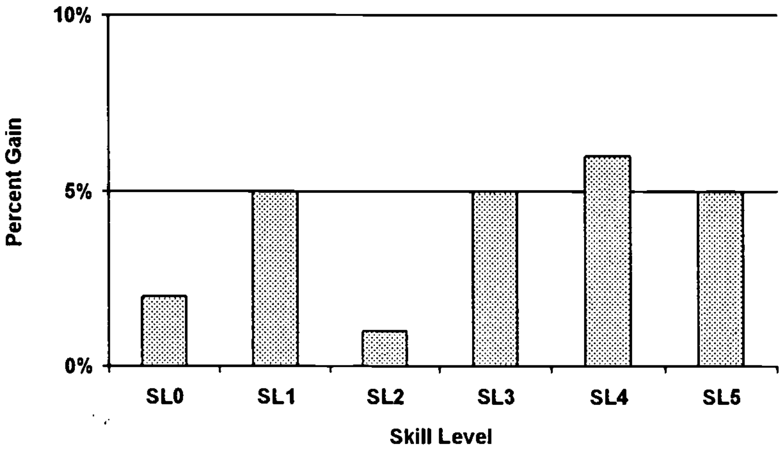
The ANCOVA results for the test of the skill level main effect were F[(5, 1540) = 55.31, p < .001]. The skill level effect accounted for about 15% of the variance while prior knowledge contributed about approximately 17% of the score variance. Between group effect sizes (adjusted for prior knowledge) were SL0 vs SL1 = .-.20, SL0 vs SL2 = .35, SL0 vs SL3 = .54, SL0 vs SL4 = .77, and SL0 vs SL5 = .71.

### 3.1.5 Conducting an Experiment

The *conduct experiment* (CE) composite was comprised of a series of questions dealing with issues like manipulating independent variables. Percent gains by SL were 2%, 5%, 1%, 5%, 6%, and 5% respectively. These values translate to pre- to post-test effect sizes of .09, .28, .05, .25, .32, and .24.

The ANCOVA F[(5, 1540) = 6.79, p < .001] revealed a small but significant effect. The covariate regression accounted for 25% of the variance while the SL main effect, by contrast only contributed about 2% to the predictiveness of the model. The contrast tests revealed that subjects had to reach SL4 in order to differ significantly control.

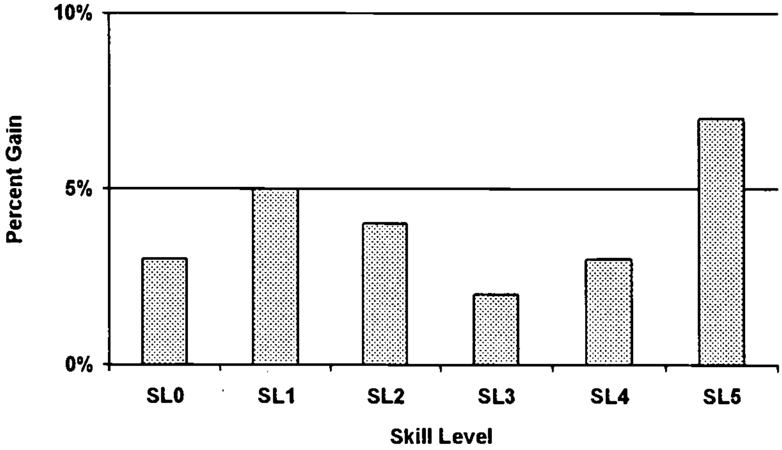
**Figure 8. CE Subscale Score Gains by SL**



### 3.1.6 Drawing a Conclusion

The *draw conclusion* (DC) score composite was formed by a series of questions asking students to draw appropriate conclusions based on textual and graphic descriptions of experimental data. Score gains were 3%, 5%, 4%, 1%, 3%, and 7%. The corresponding effect sizes were .10, .16, .13, .03, .12, and .28. Tests of the SL factor produced significant results F[(5, 1540) = 8.29, p < .001], with prior knowledge accounting for 19% of the variance. SL added approximately 3%.

**Figure 9. DC Subscale Score Gains by SL**

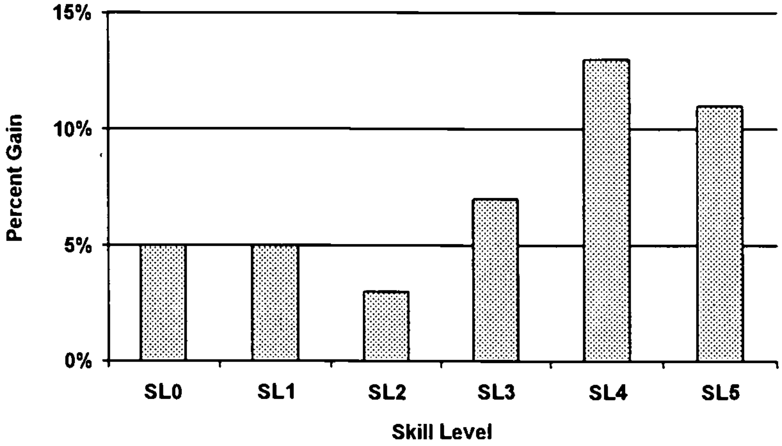


Post-hoc contrast results indicated that experimental subjects had to reach SL4 in order to outperform control on this measure. Adjusted between group effect sizes were -.08, -.01, .03, .10, and .18 for SL1-5 when compared to SL0.

### 3.1.7 Accepting or Rejecting Hypotheses

The *accept/reject hypothesis* (AH) composite score was formed from questions which simply asked testees to accept, reject, or determine inconclusiveness based on experimental data. Pre- to post-test score gains were 5%, 5%, 3%, 7%, 13%, and 11%. The corresponding within-SL effect sizes were .19, .18, .12, .30, .48, and .42.

**Figure 10. AH Subscale Score Gains by SL**

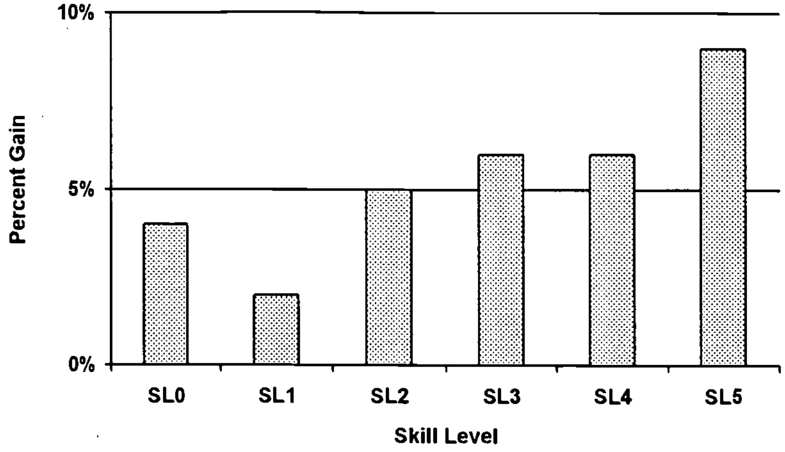


The ANCOVA results indicated that the skill level factor was significant F[(5, 1540) = 10.16, p < .001]. Overall the model accounted for about 22% of the score variance (prior knowledge captured about 18% of this variance). Post-hoc contrasts again revealed that subjects had to reach SL4 or SL5 to produce score gains which were significantly better than the control. At these skill levels, students outperformed the control by roughly 11% and 8% respectively. Adjusted between group effect sizes were -.13, -.09, .12, .42, and .37 for SL1-5 when compared to SL0.

### 3.1.8 Domain Knowledge

The domain knowledge (DK) composite consisted of a series of questions about substantive knowledge of ecological interrelationships. Pre to post-test score gains were 5%, 2%, 4%, 6%, 6%, and 9%. Within- SL effect sizes were .26, .10, .22, .35, .33, and .53 respectively.

**Figure 11. Domain Knowledge Subscale Score Gains by SL**



The overall model was significantly able to predict SL group membership F[(6, 1540) = 156.94, p < .001]. The ANCOVA indicated a significant main effect of skill level for the DK dependent variable F[(5, 1540) = 9.41, p < .001]. Here, prior knowledge contributed about 35% out of 38% of the variance accounted for by the overall model. Contrasts for DK scores indicated that students had to again reach SL4 in order to outperform control students. The adjusted between-group effect sizes (SL1-5 vs control) were -.24, -.03, .05, .16, and .37 respectively.

## 4.0 Discussion

### 4.1 The Current Test and Evaluation

The results from this large-scale, year-long test and evaluation of an intelligent tutoring system geared toward improving students' scientific inquiry skills are encouraging. Students' skills and domain knowledge grew in a linear fashion as experience with the tutoring system increased. This pre- to posttest gain was larger than for inquiry skills than for domain knowledge. Furthermore, these gains are greater for the treatment group as compared to the non-treatment control group.

While the overall results show that ISIS was instructionally more effective than traditional, teacher-based, large-class methods, the effects appear only for those students who have had a moderate amount of success with the tutor. The students who had better success using the tutor in terms of number of research assignments completed, showed higher gains from the pre- to posttest. One possible interpretation is that academically better students learned more implying that only bright students learn from computer-based forms of instruction. However, the results showed that inquiry skill level achieved predicted performance even when pretest scores were accounted for in the analysis. Furthermore, the strength of the covariate regressions suggest that initial aptitude is best predictor of final test performance, but not necessarily gains across the academic year. An alternative possibility is that there is a threshold for students to develop a coherent understanding of scientific inquiry (Schauble, et al., 1995). Students may have to progress to a certain level of experience with scientific inquiry activities in order to get "the big picture." This is supported by the design of ISIS in that students reaching the third level of skill development use the simulated biomes to conduct their experiments. It is at this point that they carry out a wide range of scientific activities: planning and conducting their experiments.

Owston (1997) asked three questions about the implementation of internet technologies in the classroom that can raised for any educational technology not just internet-based approaches. The questions address the instructional efficacy of the technology, making learning more accessible, and containing the cost of education. The outcomes of this research show that ISIS promotes learning relative to traditional, non-technology-based approaches. While the gains are not overwhelming, they are present for those students who complete a moderate number of research assignments embedded in ISIS. Providing educational opportunities for students to learn scientific inquiry skills was a point raised in our interviews with the teachers using ISIS in their classrooms. They pointed out that their students learned science-related skills faster and more systematically than in previous years. Moreover, many believed that ISIS gave students experience with scientific thinking that they could not provide in the traditional, cookbook lab settings. While anecdotal evidence must be taken with caution, the information from the teachers lends credence to the notion that technology-based learning environments, such as ISIS, provide learning opportunities that teachers cannot normally provide.

The cost of implementing ISIS is difficult to assess. Financially, the cost is not great, because the software was on loan in return for participation in the research study and it is not yet a commercial product with an associated price. The hardware used in this study was either owned by the school district or on loan, again, in return for participation in this study. The course materials were initially developed and maintained by an interdisciplinary team funded by under the research project. To continue using ISIS, though, would not be very costly. The primary cost would be to repair and upgrade the outdated hardware and to provide a technician in the computer lab. Curricular "costs" are actually more of a burden than financial costs. Teachers must decide what topics they must cut from their normal classroom curriculum in order to allocate time to using the software. In some cases, such as photosynthesis, the choice is easy, because the topic is covered in the tutoring system. However, eliminating 15-20 hours of in-class time over an academic year requires the teachers to seriously reconsider their curricular choices. These choices are not easy given the pervasive emphasis on state testing requirements.

One issue that Owston (1997) does not raise is about the implementation of technology. We have found in the FST project that teacher training is one of the more important components required for successful implementation of the tutors. Many teachers do not have experience using computers to deliver instruction to students. Instead, they use computers for recording grades and to develop paper-based handouts and exams. Teachers need not only to understand instructional software from a user's point of view, but also how to teach with it. The teacher's role in the classroom is not to put hands on the keyboard or mouse, but to stand next to the students acting as an partner in the instructional process.

The number and arrangement of computers in an educational setting is a critical issue facing the school districts. One arrangement is to have 25-30 computers networked in one room. This arrangement provides opportunities for equal access to the technology for all students.

Depending on the software, teachers most likely are able to individualized the instruction to the needs of the students. An alternative arrangement that is growing in popularity is to have 6-7 computers in the back of the regular classroom. Proponents state that this arrangement will provide teachers more opportunities to individualize the curriculum. On the other hand, some teachers in our project have expressed concern that the computers in the classroom arrangement will cause classroom management problems. They are concerned that teachers will not be able to provide equal access to the technology. If access is based on student-choice, then students who are less computer literate or hold negative attitudes towards computers may have benefit as much as computer enthusiasts. If access is teacher-driven, administering access time and curriculum covered may burden the teachers with additional administrative workload. Another issue is the amount of time the computers would be used. In our lab-based arrangement, computers are used almost constantly. Skeptics of the classroom-based arrangement, fear that the computers in the classroom will not be used as frequently by the students lowering the total access time to the available technology.

## Summary

The Fundamental Skills Training (FST) Project is a multi-year research project designed to develop, implement and evaluate three intelligent tutoring systems in public education settings. The study presented here is the first year large-scale field evaluation of ISIS, an intelligent tutoring system designed to teach scientific inquiry skills to high school students. ISIS was found to improve student's scientific inquiry skills more than traditional, large-class instruction. Changes to the design of ISIS have been made based on the evaluation, teacher monthly reports, and extended interviews with teachers and students.

# References

Campbell, D. T. & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research.* Rand McNally College Publishing Company: Chicago.

Carey, S. & Smith, C. (1993). *On understanding the nature of scientific knowledge*. Educational Psychologist, 28(3), 235-251.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L.B. Resnick (ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum.

Collins, A., Hawkins, J., & Carver, S.M. (1991). A cognitive apprenticeship for disadvantaged students. In B. Means, C. Chelemer, and M.S. Knapp (Eds.), *Teaching advanced skills to at-risk students: Views from research and practice*. San Francisco: Josey-Bass.

Glynn, S. M. & Duit, R. (1995). Learning science meaningfully: Constructing conceptual models. In *Learning science in the schools: Research reforming practice*. Eds. Glynn, S. M. & Duit, R. Mahwah, NJ: Lawrence Erlbaum Associates.

Hsieh, P. Y., Miller, T. M., Hicks, Lorenz, K.P. (1993). *Identification of effective teaching behaviors.* (Technical Report AL/HR-TR-1993-0091). Brooks AFB, TX.

Jonassen, D. H. (1991). *Objectivism versus constructivism: Do we need a new philosophical paradigm?* Educational Technology Research and Development, 39(3), 5-14.

Jonassen, D. H. (1991a, Sept). *Evaluating constructivistic learning*. Educational Technology.

Kulik, C. L. & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior*, 7, 75-94.

Kulik, C. L., Kulik, J. A., & (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior*, 1, 59-74.

Kulik, J. A., Bangert, R. L., & Williams, G. W. (1991). Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology*, 75(1), 19-26.

National Research Council (1996). *National Science Education Standards* .Washington, D.C.: National Academy Press.

Owston, R. (1997). The World Wide Web: A technology to enhance teaching and learning? *Educational Researcher*, 26 (2), 27 - 33.

Rutherford, F.J. & Ahlgren, A. (1990). *Science for all Americans*. New York: Oxford University Press.

Schauble, L., Glaser, R., Duschl, R.A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, 4(2), 131-166.

Yager, R. E. (1995). Constructivism and the learning of science. In *Learning science in the schools: Research reforming practice*. Eds. Glynn, S. M. & Duit, R. Mahwah, NJ: Lawrence Erlbaum Associates.

SE060483

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*Educational Resources Information Center (ERIC)*

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

Title: Evaluation of an Authentic Learning Environment for Teaching Scientific Inquiry Skills

Author(s): Steuck, K. & Miller, T.

| | |
|---|---|
| Corporate Source: Armstrong Laboratory & Mei Technology Corp. | Publication Date: 3/97 |

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

| | The sample sticker shown below will be affixed to all Level 1 documents | The sample sticker shown below will be affixed to all Level 2 documents | |
|---|---|---|---|
| ☒<br>**Check here**<br>**For Level 1 Release:**<br>Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy. | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY<br>Sample<br>TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY<br>Sample<br>TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | ☐<br>**Check here**<br>**For Level 2 Release:**<br>Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy. |
| | **Level 1** | **Level 2** | |

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

*"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."*

**Sign here→ please**

| | | |
|---|---|---|
| Signature: | Printed Name/Position/Title: Kurt Steuck | |
| | Telephone: 210 536 2034 | FAX: 210 536 2902 |
| | E-Mail Address: steuck@alhrt.brooks.af.mil | Date: 5/21/97 |

Dr. Kurt Steuck
7909 Lindbergh Dr.
Brooks AFB TX 78235-5352

CUA

**THE CATHOLIC UNIVERSITY OF AMERICA**
*Department of Education, O'Boyle Hall*
*Washington, DC 20064*

*800 464-3742 (Go4-ERIC)*

April 25, 1997

Dear AERA Presenter,

Hopefully, the convention was a productive and rewarding event. We feel you have a responsibility to make your paper readily available. If you haven't done so already, please submit copies of your papers for consideration for inclusion in the ERIC database. If you have submitted your paper, you can track its progress at http://ericae2.educ.cua.edu.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are soliciting all the AERA Conference papers and will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and stet **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can mail your paper to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:
AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/E