

Assessing Dynamics in Computer-Based Instruction

John F. Pane

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213 USA
+1 412 268 1449
pane+chi96@cs.cmu.edu

Albert T. Corbett

HCI Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
+1 412 268 8808
corbett@cmu.edu

Bonnie E. John

Computer Science Department,
Psychology, and HCI Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
+1 412 268 7182
bej+@cs.cmu.edu

ABSTRACT

We present an evaluation of a multimedia educational software system that includes text, graphics, animations, and simulations. When compared with an informationally equivalent control environment that used text and carefully selected still images, we found little evidence that the dynamic presentations enhanced student understanding of the declarative information in this lesson. Furthermore, students cannot be relied on to take full advantage of exploratory opportunities in computer-based instruction. These results prescribe further investigation of whether and how computer-based multimedia can be used effectively in education and training.

Keywords

animation, simulation, multimedia, computer-based learning.

INTRODUCTION

Multimedia educational environments are rapidly growing in popularity. Journals and international conferences devoted to educational multimedia have sprung up, multimedia encyclopedias are a popular throw-in with personal computer hardware, and as computing costs decline multimedia learning environments are destined to make substantial inroads into schools at all levels. It is plausibly assumed that the interactive nature of multimedia environments along with video and audio presentations will engage the students' interest and intellect, particularly contemporary students who have grown up with television and video games. However, relatively little is known about the conditions under which multimedia presentations are effective.

In this paper we examine the impact of computer-based movies and simulations on students' understanding of time-varying biological processes. A variety of studies have

examined the use of such animations in computer-based education. A range of results has been obtained, but many studies report a positive impact of animations on learning and student motivation. In most, but not all, of these studies, the computer-based animation condition is compared to a control condition in which students either do nothing, just read text or just examine still graphics. While these outcomes are of practical importance in education, they do not directly bear on the genuine value-added of dynamic computer-based presentations on learning unless a principled effort is made to develop an informationally equivalent control condition. In this paper we focus on the process of developing an informationally equivalent control condition in assessing the value added by dynamic presentations.

THE ACSE ENVIRONMENT

This study employs the Advanced Computing for Science Education (ACSE) environment [12]. The ACSE environment focuses on the understanding of complex time-varying processes. It is designed to afford students the opportunity to actively design and execute experimental investigations of the mechanisms underlying these dynamic processes and to be a living document of the student's investigation and growing understanding.

The ACSE software provides a multimedia document containing text, still graphics, movies, and simulations. The system provides the student with tools for navigating through the lesson, viewing the movies, and manipulating and running the simulations. Movies include any dynamic presentations that are the same each time they are viewed, e.g., animations, videos, and time-lapse photography. Simulations include any graphics, tables or other output generated under programmatic control of the student. Simulation outcomes are subject to changes the student makes, as well as probabilistic variations, and thus may be different each time they are viewed.

In this study we are interested in the impact of both movies and simulations on students' understanding of declarative facts about time-varying biological processes. As described in the next section, a major thrust of this study was to

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

CHI 96 Vancouver, BC Canada

© 1996 ACM 0-89791-777-4/96/04..\$3.50

design a formative evaluation that would reveal the real value added by dynamic presentations. Much of the effort went into developing an informationally equivalent control condition that relied exclusively on static images and text.

Other research has investigated the effectiveness of movies and simulations for declarative knowledge acquisition, with mixed results. Studies investigating the use of movies to teach mathematics and dental hygiene have found positive effects [1, 4]. In addition, studies using simulations in biology and economics found positive effects [3, 6]. But these results are balanced by similar studies in content areas such as physics and computer science where no significant improvements are found [5, 13, 14, 15]. Overall, none of the above studies took a disciplined approach to information equivalence between the experimental and control conditions.

Some studies are notable for the care with which they constructed equivalent comparison conditions. One study found an advantage of movies for long-term recall of episodic story structure, although not for immediate recall [2]. Here the comparison condition was a text-only condition that controlled for episodic story structure rather than information equivalence. Other studies have found value-added for movies in acquiring procedural knowledge, where movies improved immediate recall, but not long term recall, of how to use a direct-manipulation computer interface [8, 9, 10]. It is questionable whether procedural results are applicable to the acquisition of declarative knowledge in any case, but again, these studies employed a text-only comparison condition.

We don't doubt that visual images can convey information that is difficult to capture in text. In contrast, this study examines the relative impact of dynamic visual presentations on the acquisition of declarative knowledge. We compare the impact of two informationally equivalent environments, one that combines text, still graphics and dynamic presentations, with a second that uses only text and still graphics.

FORMATIVE EVALUATION DESIGN

Development and evaluation of educational software depends on a close collaboration between software developers and instructors. The ACSE project is a collaboration among the following Carnegie Mellon departments: Computer Science, Human Computer Interaction Institute, Biological Sciences, and Center for Light Microscope Imaging and Biotechnology. The target domain is an upper level undergraduate course in developmental biology and the course instructor identified two topics involving dynamic processes that he viewed as difficult to teach.

The first lesson is titled *Sea Urchin Gastrulation* and examines a process common to many organisms in early development in which cells adopt specialized roles and migrate to appropriate locations in the embryo. The second lesson is titled *The Early Development of Drosophila Melanogaster* and examines the way that gradients of

molecules produced by the mother and stored within the fertilized egg can result in differences among embryonic cells and the generation of patterns in the organism. Each lesson is about 50 pages long, although these pages correspond to the size of the content window on a 13-inch computer screen, and thus have less information than a typical textbook page. The lessons contain multiple high-resolution light microscope images and movies, as well as multiple simulations.

Review questions are incorporated in each lesson and serve as the dependent measure of student learning. Three classes of review questions are distinguished, based on the form in which relevant information is presented in the ACSE condition. The review questions tapped information that was presented primarily in the form of (1) text and static graphics, (2) movies, or (3) simulations. In the control condition all of the questions were based on information that was presented in form of text and static graphics.

Control Condition

The most challenging issue in evaluating multimedia learning environments is the construction of an appropriate comparison condition. The goal in this condition is to convey the information presented in the movies and simulations as well as possible with still images and text. In our ACSE evaluation, we adopted the model of a conventional biology textbook for the comparison condition. We developed a *textbook* variation of the ACSE environment in order to reduce the possible confounding effects of comparing the computer with a paper-based control condition. We used the following guidelines in constructing the contents of the control condition.

Select Key Frames

One great challenge in setting up the textbook comparison condition was selecting the appropriate still images. The lesson designer was interviewed to determine the replacement elements for the control condition. For each movie and simulation element, we asked him to reflect on the pedagogical goal of the element by asking questions like: What is the student expected to learn from this lesson element? What are the important events in this time-varying process? Which still images best represent these important events? What captions are required? How would a textbook best describe this material? Videotapes of these sessions were inspected to create the following strategy: choose frames representing the initial and final conditions, and intermediate frames that represent important incidents. However, in the event that there is a series of simulations with identical initial conditions, it is not necessary to repeat the frame representing the initial condition each time.

Choose Illustrative Simulation Runs

In the ACSE lesson, when students encounter a simulation, they are directed to run it without modification. Then they are instructed to make a specific modification to the simulation and to run it again. After these directed runs, they are encouraged to make further modifications to explore the behavior of the system. In deriving the control

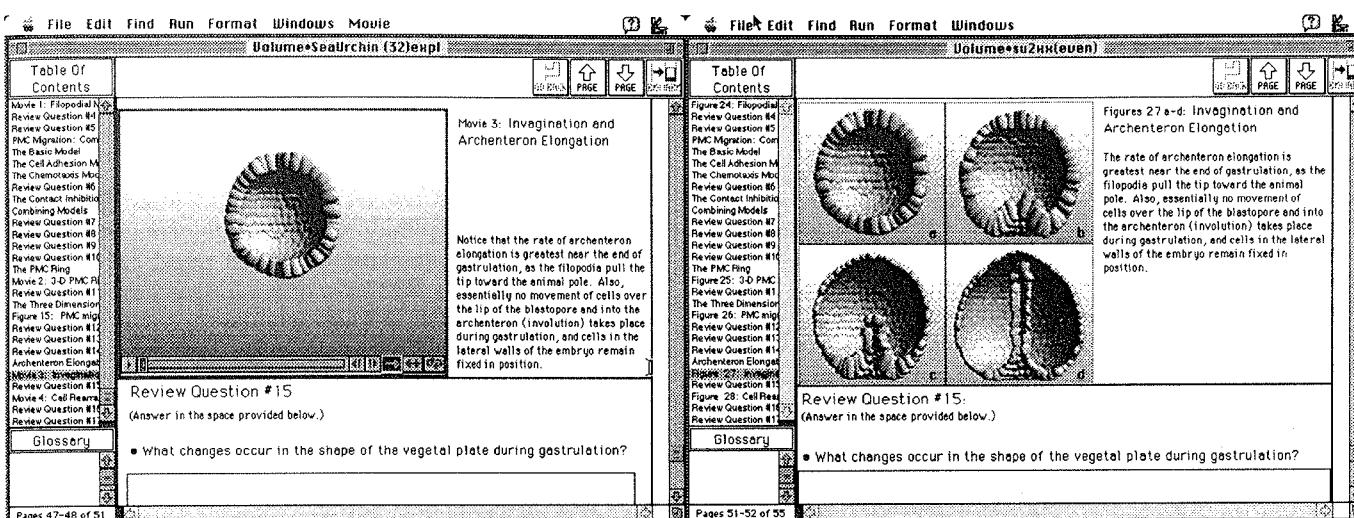


Figure 1: (a) ACSE lesson containing a movie. (b) Equivalent lesson with text and graphics.

lesson, each of the specified simulation runs is represented by still images. It is essential that these actually illustrate the necessary dynamics, so that the review question is answerable in the control condition.

Locality and Economy of Space

Biology textbooks typically have space limitations and a survey of representative books revealed that they seldom devote more than two still pictures to a particular topic. Furthermore, it is desirable to maintain locality of information. We limited the number of still pictures that replace each movie and simulation to a maximum of four, occupying a maximum of two lesson pages. The average number of still pictures representing each movie or simulation was two.

Preserve Visual Quality

Since visual quality may affect the information a student can get from an image, we prohibit any reduction in the number of pixels per image (spatial resolution) or number of bits/pixel (pixel resolution) between the images used in the movies or simulations and their still-graphic counterparts.

Replace Program Code with Text

ACSE shows small pieces of the Pascal simulation programs to the student. In some instances, this Pascal code is displayed in a natural language form, but in other cases the syntax of the programming language shows through. In the control lesson, simulation code is always translated into natural language or simple mathematical formulas.

Explanatory Text

In deriving the control elements for movies and simulations, it is often necessary to supplement the selected still images with explanatory text that describes some behavior that is observable in the movie or simulation, or to fill in the gaps between the chosen stills. This text was provided in both the ACSE and the control lessons.

Figure 1 displays the two versions of one page of the developmental biology lesson about sea urchin gastrulation. Figure 1(a) displays an animation of a stage of development in the sea urchin embryo. The student clicks the play button to watch the process unfold. Figure 1(b) displays the equivalent still images. Students can scroll through the lesson pages with the scroll bar, or page up and down with the buttons located in the upper right corner of the window. The Table of Contents allows students to directly access major sections in the text.

HYPOTHESIS ON MEDIA EFFECTIVENESS

Our interviews of the instructor on curriculum goals and a careful inspection of the review questions led us to formulate a hypothesis concerning the effective deployment of educational multimedia. We informally observed two distinct instructional goals: some review questions tended to focus on understanding processes, while other review questions tended to focus on end products. An example of the former is, "Describe the appearance and motile behavior of these cell processes," while an example of the latter is, "At the end of active migration, what is the arrangement of [cells] in the embryo?" Our tentative hypothesis is that dynamic presentations are more effective to portray processes, while still images are sufficient to portray end products.

THE EXPERIMENT

Participants

Thirty four Carnegie Mellon students (13 male, 21 female) taking an upper level developmental biology class participated in the study. Performance in the lab session constituted 10% of their course grade.

This set of students was divided into two groups of seventeen through a matching process. Students were rank ordered on the basis of prior test performance in the course, with QPA used to break ties. Within successive pairs in this list, one student was randomly assigned to group A and the other to group B. The two groups were well matched on

class performance (68.0 vs. 68.0), QPA (3.08 vs. 3.14) and sex (10 vs. 11 females). All except one student (group A) had completed or placed out of at least one university-level computer programming course.

Materials

All students worked through the two ACSE lessons. The first lesson, entitled *Sea Urchin Gastrulation*, examines a process in early development that is common to many organisms, where cells adopt certain roles and migrate to appropriate locations in the embryo in order to fulfill those roles. It is 51 pages long, and in addition to text contains 15 images and figures, 4 movies, 18 fragments of simulation code drawn from 6 simulations, and 17 review questions. The second lesson, entitled *The Early Development of Drosophila Melanogaster*, examines the way that gradients of molecules produced by the mother and stored within the unfertilized egg can result in differences among embryonic cells and the generation of patterns in the organism. It is 55 pages long, and in addition to the text contains 23 images and figures, 3 movies, 10 fragments of simulation code drawn from 7 simulations, and 10 review questions.

During the time span between the first and second lessons, a table of contents feature was added to the software to assist the student in navigating and tracking progress through the lesson. This new feature was present for lesson 2 for both the ACSE and control conditions. In addition, in order to foster greater use of the simulations in lesson 2, we focused most of the review questions on the processes portrayed in the simulations, and introduced explicit recommendations to run the simulations along with suggested manipulations to perform between runs.

Each review question was sorted into one of three categories, by determining the lesson element that was most relevant to answering the question: text or still graphics; movies; or simulations. The 17 review questions in the first lesson fall into the following categories. Seven questions focused on the initial material in the lesson that contained no movies or simulations. These questions serve as a control to ensure the equivalence of the two student groups. Five questions focused on the movies and five focused on the simulations. In the second lesson nine review questions focused on the simulations while the tenth question focused on a movie. The movie and simulation questions were further classified depending on whether they focused on dynamic processes or end results. A biology graduate student performed this classification, which is shown in Table 1.

	Lesson 1	Lesson 2
Movie	4P / 1E	0P / 1E
Simulation	4P / 1E	3P / 6E

Table 1: Classification of review questions into *process* or *end result*. (P=process; E=end result).

After finishing work on each lesson, participants were given a short paper questionnaire. It asked about prior

computer experience and satisfaction with the lesson, as well as opinions about the best and worst aspects of the system and suggestions for improvements.

Design

The study was conducted during two evening lab sessions at appropriate points in the course curriculum. The first lesson was presented midway through the course, and the second lesson was presented in the last week. All students worked through the first lesson in the first session, with group A in the dynamic ACSE condition and group B in the static control condition. In the second session, all students worked through the second lesson, with group B in the dynamic ACSE condition and group A in the static control condition.

Procedure

Each session was held in a large computer laboratory that is partitioned into two sections. All of the ACSE participants were assigned to one half of the lab; and all of the control participants were assigned to the other. Participants received a one-page instruction sheet about the features of the ACSE environment. A proctor directed them to a computer which had the appropriate lesson opened to page one and recorded the starting time. Participants worked on identically configured Macintosh IIxi computers with 9 MB of memory and 13-inch 8-bit color monitors.

Students had no *a priori* knowledge that there were two different lesson formats. Students were permitted to work as long as they wished, were informed that their answers to the review questions would be graded, and that each lesson was worth 5% of their course grade. There was the additional motivation that the material might appear on an exam. The lessons were not made available to the students for study outside the laboratory session.

Feature usage and user inputs were timestamped and automatically logged by the software. When participants indicated that they were finished, the time was recorded by the proctor, and they were given the questionnaire.

Due to scheduling conflicts, there were six students who attended a makeup session for the first lesson (four ACSE; two control) and two students who attended a makeup session for the second lesson (one ACSE; one control). The procedure used during these makeup sessions was identical to the main session. The only known confounding factor is the possibility that these participants discussed the lessons with their peers before attending the makeup session. Three students who participated in the first session did not participate in the second session (second session group assignments: one ACSE, two control).

The answers to review questions were scored by an independent grader, a biology graduate student, who was blind to condition. The grader was not familiar with the details of this study. In a debriefing session after the experiment concluded, she reported nothing that would have biased the grading.

QUANTITATIVE PERFORMANCE RESULTS

Pretest Measures as Performance Predictors

We found no significant correlation between the pretest measures (QPA, Prior Exams) and performance overall or performance on any of the categories of questions.

Time on Task

On the first lesson, the ACSE group took 146 min ($sd=36$) to finish, while the control group took 101 min ($sd=23$). This difference was significant [$t(32)=4.4$, $p<.001$]. On the second lesson, the ACSE group took 130 min ($sd=32$) to finish, while the control group took 100 min ($sd=35$). This difference was significant [$t(28)=2.5$, $p<.05$].

Performance on Review Questions

We measured the students' scores on the four categories of questions based on (a) identical static text and still images at the beginning of lesson 1, (b) lesson 1 movies, (c) lesson 1 simulations and (d) lesson 2 simulations.¹ Recall that the movie and simulation questions in lesson 1 focus overwhelmingly on dynamic processes while the simulation questions in lesson 2 focus mostly on end-products. The results are shown in Table 2. The numbers represent the average score on questions in the categories listed, expressed as a percentage of the possible points in that category. A significant difference between groups was found only on the lesson 1 simulation questions. There the ACSE group scored about 10% higher than the control group [$t(32)=2.153$, $p<.05$].

	Lesson 1		Lesson 2	
	Static (7)	Movie (5)	Simulation (5)	Simulation (9)
ACSE	72 ($sd=10$)	70 ($sd=12$)	78 ($sd=10$)	64 ($sd=12$)
Control	71 ($sd=10$)	72 ($sd=14$)	71 ($sd=9$)	68 ($sd=17$)

Table 2: Performance on review questions in percent correct by question type (# of questions in parenthesis).

The lesson 1 simulation questions focused heavily on dynamic processes rather than end products. To determine if this distinction accounts for the significant effect in Table 2, we combined the simulation questions for lessons 1 and 2, and performed a pairwise t-test comparing the process-oriented questions and product-oriented questions. This test was non-significant.

STUDENT ATTITUDES

There were no significant differences between the ACSE condition and control condition on any of the survey questions, although the responses were generally favorable.

¹ On Lesson 2, one participant visited only 34 of the 55 lesson pages, ran only three of the seven simulations, and answered only two of the ten review questions. On the questionnaire, this participant reported that "I didn't have time to do it because of other work (tests)." This participant was removed from the analysis of lesson 2.

Further information about the survey responses can be found in [11].

FORMATIVE USABILITY EVALUATION

We collected both direct feedback from the students with the questionnaire, and software logs containing records of usage patterns and problems. These provide insight into the usability of the ACSE environment which we can use to improve future versions of the system.

Lesson 1

Students made far less use of the movies and simulations than we anticipated. Students ran each movie an average of 1.4 times and ran each simulation an average of 1.3 times.

Three types of student comments stood out after the session. First, students praised the high quality of the graphics. Second, they complained that the system ran slowly, particularly the simulations. Third, they complained that there was no table of contents to assist them in finding relevant information when answering the review questions.

Lesson 2

As described above, we introduced two changes for lesson 2. First, a table of contents was provided to make it easier for the student to navigate through the environment. Secondly, we focused the review questions on the simulations and added explicit recommendations to run the simulations along with suggested manipulations to perform between runs. These relatively small changes in the environment had a small positive impact on usage of the ACSE features. While the average number of movie runs remained essentially unchanged at 1.3, the average number of runs per simulation increased to 2.0. Software logs show that the table of contents was heavily used and three students reported that this feature was the best thing about their experience with lesson 2.

GENERAL DISCUSSION OF PERFORMANCE

We draw several important conclusions from this evaluation of the dynamic presentations in a computer-based environment. First, dynamic presentations are not a panacea for instructional difficulties. With well-defined instructional goals, appropriate learning data, and an informationally equivalent control environment with text and carefully selected still images, we found little evidence that dynamic presentations enhance student understanding. Moreover, student attitudes concerning the dynamic and static conditions did not vary significantly.

Second, even motivated students cannot be relied on to take full advantage of exploratory opportunities. The upper level biology class was largely populated by biology majors at a selective university. This group of students reflects about the highest general motivation level that can be realistically expected in a K-12 or university environment. Nevertheless, we were disappointed that students made relatively little use of the potential to explore the experimental manipulations afforded by the simulations. Students ran an average of 1.3 simulations in the first lesson. When two specific simulations were recommended

in the second lesson, students averaged exactly 2.0 simulation runs. This suggests that substantially more guidance must be built into such experimentation environments.

Comprehension Differences

Despite these general caveats, dynamic presentations did enhance student comprehension in lesson 1 simulation questions and this is a moderately large, 0.7 standard deviation effect. Further analyses indicated that this effect did not depend on process-oriented questions, contrary to our hypothesis.

However, the lesson 1 simulation questions did have a unique property. This was the one situation where students in the dynamic condition could receive more visual information than was present in the control condition. Each time students ran the sea urchin simulations in lesson 1, they observed a different probabilistically varying development sequence (in addition to the set of static images that were presented in both the dynamic and static environments). This was not true of the movies, which always display exactly the same dynamic sequence, or the lesson 2 simulations which did not have probabilistic behavior. In short, the lesson 1 simulation condition presented more visual information than the corresponding control condition. This suggests that dynamic presentations will be useful when information concerning process variability can be gleaned from multiple runs – more runs than can be reasonably portrayed in a set of still images.

It should be emphasized that these conclusions are limited to dynamic presentation of declarative knowledge. We have no reason to expect that they generalize to the acquisition of procedural knowledge.

The magnitude of the performance improvement should be considered in the context of the small amount of time allotted to each lesson. We do not believe the minimal effect obtained in this study generalizes to a situation in which students are actively engaged in experimentation and modifying simulation parameters. The challenge is to engage students in this activity; it does not come readily even to students who are well prepared and relatively motivated. The tools themselves are complex as are the reasoning patterns they support. Students received no instruction on the features and effective use of the system. Instead, students jumped right into using the system with difficult topics that normally arise near the end of the course. It may be necessary to scaffold effective feature usage with more modest topics introduced right from the beginning of the course. In addition, the teacher may need to demonstrate the features of the system through routine use in the classroom lectures before students will adopt and benefit from advanced capabilities [7].

It may also be the case that even the control lesson in this study was superior to the other teaching materials that are available to the instructor. It is derived from a carefully constructed lesson that covers exactly the topics desired, in exactly the order that this particular instructor would

choose to present them. It uses still graphics that were generated by a simulation which would otherwise not have been available were it not for the ACSE environment. This is analogous to writing a custom textbook for the course and instructor. This spin off of the careful formative evaluation is of educational significance, although it does not bear directly on human computer interaction issues.

Time Differences

In both lessons, the control lesson is approximately the same length as the ACSE lesson. However, the process used to derive the control ensures that it has no more total information than the ACSE lesson. This, in combination with the fact that each ACSE simulation takes several minutes to run and can be run an unlimited number of times, suggests that the ACSE group will take more time to work through the material. This is confirmed by our observations of total time.

While much of the additional time used by the ACSE group is spent in running simulations, we found no significant correlation between the number of simulations run by the student and their performance on those review questions that were directly related to those simulations.

Analysis of the first 20 pages of lesson 1, which are the same between groups, supports the expectation that any time or performance difference between groups is due to the experimental manipulation that appears later in the lesson. Participants might have spent more time on a lesson if they were enjoying the experience or felt that it was a productive use of their time. Alternatively, participants might have felt some obligation to complete the lesson, and thus would have been less satisfied if it was not enjoyable or seemed an unproductive use of time. Since student satisfaction measures do not show significant correlations to time spent, neither of these speculations can be confirmed.

USABILITY RESULTS AND DISCUSSION

Navigation

As discussed earlier, the table of contents feature proved essential in allowing students to effectively navigate through the environment and in shaping their overall evaluation of the environment.

However, not every page of the lesson has a table of contents entry, so some other method of scrolling must be used to reach those pages. One student exclusively used the table of contents for navigation, thus skipping pages of the lesson that do not appear there. This happened even though (a) this student had used the other scrolling mechanisms on lesson 1, (b) there is a page number indicator visible at all times, and (c) the text had several cross references to these pages. This student's behavior surprised us and suggested that page numbers next to the table of contents entries might alleviate this problem.

Management of Information

There were six comments (one ACSE, five control) about difficulty in remembering material from page to page. The

logs showed that when students are answering the review questions, they often scroll back through the material multiple times in the midst of typing their answers. This problem may be relieved when previously-planned features of ACSE – a glossary, bookmarks and live cross-reference links – are completed. One subject suggested that the system provide a scratchpad that the user can use to record notes and simulation results.

Modifying Simulations

In ACSE, we experimented with using natural language syntax for presenting simulation code. One example is a transformation of Pascal procedure calls, where:

`Regulator(bicoid, 1000, activates, hunchback, 50);`

becomes:

Regulator: when the concentration of **bicoid** is greater than **1000** units, it **activates** the expression of **hunchback** with an efficiency of **50%**.

This special formatting posed some problems for students when they attempted to edit the statement, because the underlying Pascal syntax appears at that time and changes must be entered using that syntax. The natural language syntax reappears when editing is completed. The logs show that this transition between representations was a problem for many students: they encountered syntax errors and sometimes inadvertently destroyed the annotations that provide the natural language syntax. We conclude that while this implementation was easiest for the development of the ACSE software, we must change the editor so students can work exclusively in the natural language syntax.

We also observed students having difficulty with the Pascal syntax in other cases: a) real numbers must have a digit before the decimal place, so values like **.8** are not accepted; and b) commas are not permitted in large numbers, so **32,000** must be entered as **32000**. These problems highlight the limitations of using a programming language in the context of natural language. Buttons and sliders, which were previously-planned but not yet implemented features of the software design, can be used to provide a direct-manipulation interface to numeric parameters. The system could also have been improved to selectively allow deviation from the Pascal standard, to permit incorrect but unambiguous syntax to be interpreted in the expected way.

Workstation Requirements

The predominant positive comment from both groups concerned image quality (22 comments from the ACSE group, and 20 from the control group). Of these, 10 of the ACSE comments specifically mentioned movies and simulations. The lessons make extensive use of large, high resolution color graphics. It would be inappropriate to attempt to use these lessons on systems that do not support color graphics or that have miniature screens. The memory and processing demands of these high-quality graphics

were at the upper limit of the capabilities of the systems that were used in this study.

The most common negative comment regarded speed, and this was biased heavily toward users of the ACSE lesson. Forty ACSE comments and 7 control comments indicated that the system was too slow. In addition, on lesson 2, three students who had been in the ACSE condition for lesson 1 commented that the best thing about lesson 2 (when they were in the control group) is that it was faster than lesson 1. This can be explained by the processing requirements of the simulations, which are present only in the ACSE condition. In many cases the simulations took several minutes to run on the participants' computers. This indicates that the system was underpowered for this task. Clearly it would be unwise to attempt to use these fully featured lessons on less powerful machines. However, despite this performance, students seemed to like using the simulations, as nine students who had been in the ACSE group for lesson 1 complained about missing the interactivity when placed into the control group for lesson 2.

CONCLUSIONS

While the usability problems listed above may have had a negative impact on student performance, this evaluation suggests that designers of educational software should exercise caution in their use of multimedia in computer-based instruction. The mixed results of the comprehension and time measures dictate that we look carefully at the underlying pedagogy of science education. Merely using animation and simulation capabilities of modern computers does not guarantee improvement in students' learning. Well designed static graphics and text may be just as effective, and much cheaper to produce and use, than animations and simulations.

We need to find ways to measure investigative thinking more objectively, so that we can test multimedia learning environments in many educational situations. Perhaps movies and simulations are more motivating, so students will spontaneously spend more time with them even outside a structured laboratory setting, and thereby learn more. Perhaps longer-term usage of the ACSE environment for homework assignments or routine use of ACSE in classroom lectures would magnify the performance benefits. Or, these dynamic media may be more effective with student populations different than the high achievement group examined in this study. In short, much more research needs to be done to understand the requirements in lesson content and presentation media to realize the educational potential of computer-based learning environments. The method used in this study to carefully construct an informationally equivalent control condition can serve as a guide to this research.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grant Number MDR-9150211. Bonnie John's participation is supported by the Advanced Research

Projects Agency, DoD, and monitored by the Office of Naval Research under contract N00014-93-1-0934. Al Corbett's participation is supported by the Office of Naval Research Grant Number N00014-91-J-1597. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NSF, ARPA, ONR, or the U.S. Government. The authors acknowledge Laurie Damianos, Chuck Ettensohn, Kirsten Guss, Ken Koedinger, Glenn Meter, Phil Miller, and Scott Vorthmann for their contributions to the ACSE project and this evaluation.

REFERENCES

1. Baek, Y. K. and Layne, B. H. Color, Graphics, and Animation in a Computer-Assisted Learning Tutorial Lesson. *Journal of Computer-Based Instruction*, 15, 4 (Autumn 1988), 131-135.
2. Baggett, P. Structurally Equivalent Stories in Movie and Text and the Effect of the Medium on Recall. *Journal of Verbal Learning and Verbal Behavior*, 18 (1979), 333-356.
3. Grimes, P. W. and Willey, T. E. The Effectiveness of Microcomputer Simulations in the Principles of Economics Course. *Computers & Education*, 14, 1 (1990), 81-86.
4. Lautar-Lemay, C. Comparison of Computer-Assisted-Learning With Traditional Lecture and Reading Assignment. In *Proceedings of the International Conference on Computer Assisted Learning in Post-Secondary Education*. (Calgary, Canada, May 1987), 293-294.
5. Lawrence, A. W., Badre, A. N., and Stasko, J. T. Empirically Evaluating the Use of Animations to Teach Algorithms. Georgia Institute of Technology, Technical Report GIT-GVU-94-07, (Atlanta, 1994).
6. Lazarowitz, R. and Huppert, J. Science Process Skills of 10th-Grade Biology Students in a Computer-Assisted Learning Setting. *Journal of Research on Computing in Education*, 25, 3 (Spring 1993), 366-382.
7. Miller, P., Pane, J., Meter, G., and Vorthmann, S. Evolution of Novice Programming Environments: The Structure Editors of Carnegie Mellon University. *Interactive Learning Environments*, 4, 2 (1994), 140-158.
8. Palmiter, S. and Elkerton, J. An Evaluation of Animated Demonstrations for Learning Computer-based Tasks. In *Proceedings of the ACM CHI'91 Conference on Human Factors in Computing Systems*. (New Orleans, April 1991), 257-263.
9. Palmiter, S. and Elkerton, J. Animated Demonstrations for Learning Procedural Computer-Based Tasks. *Human-Computer Interaction*, 8 (1993), 193-216.
10. Palmiter, S., Elkerton, J., and Baggett, P. Animated Demonstrations vs Written Instructions for Learning Procedural Tasks: a Preliminary Investigation. *International Journal of Man-Machine Studies*, 34, 5 (1991), 687-701.
11. Pane, J. F. Assessment of the ACSE Science Learning Environment and the Impact of Movies and Simulations. Carnegie Mellon University, School of Computer Science Technical Report CMU-CS-94-162, (Pittsburgh, PA, June 1994).
12. Pane, J. F. and Miller, P. L. The ACSE Multimedia Science Learning Environment. In *Proceedings of the 1993 International Conference on Computers in Education*, T.-W. Chan, Ed. (Taipei, Taiwan, December 1993), 168-173.
13. Rieber, L. P. The Effects of Computer Animated Elaboration Strategies and Practice on Factual and Application Learning in an Elementary Science Lesson. *Journal of Educational Computing Research*, 5, 4 (1989), 431-444.
14. Rieber, L. P., Boyce, M. J., and Assad, C. The Effects of Computer Animation on Adult Learning and Retrieval Tasks. *Journal of Computer-Based Instruction*, 17, 2 (Spring 1990), 46-52.
15. Stasko, J., Badre, A., and Lewis, C. Do Algorithm Animations Assist Learning? An Empirical Study and Analysis. In *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems*. (Amsterdam, April 1993), 61-66.