

Assessing the effectiveness of a multimedia-based lab for upper division psychology students

ADRIENNE Y. LEE, DOUGLAS J. GILLAN, and CHARLES L. HARRISON
New Mexico State University, Las Cruces, New Mexico

The efficacy of multimedia-based training in producing increased learning was evaluated. Two multimedia software packages were compared with live or videotaped lectures on the same material. Results differed by type of student (low or high initial knowledge) and type of program. Multimedia training programs may need to provide more aid to students with less initial knowledge in the training domain. In addition, the effects of multimedia may be subtle and therefore require rigorous evaluation.

Computers can be used to provide students with an opportunity to actively learn material and perform tasks in the same way as professionals in the field. Students' classroom learning in psychology could be enriched if a full computer laboratory to augment psychology courses could be developed. Currently at New Mexico State University, only one method class is available for upper division students, and there are no labs connected with upper division classes; however, some courses have a special need for labs (e.g., teaching students in a learning course operant conditioning using live rats). Although the initial cost may be high, it is more cost efficient for long-term laboratory use to develop a computer lab. For example, the estimated cost for purchasing and maintaining laboratory rats is \$2,000 per student (Graham, Alloway, & Krames, 1994); however, using a computer virtual rat has an initial cost of less than \$2,000 for a single computer and software, and more than one student can use the virtual rat software in a computer lab over the course of a semester. Thus, the availability of inexpensive personal computers has led to the possibility of developing computer software dedicated to training and tutoring students (Lee, 1992; Rosen & Petty, 1992; Welsh & Null, 1991). In addition, computer-generated animation and multimedia capabilities of computers provide an opportunity to expand the ways in which students can learn (Hapeshi & Jones, 1992; Mayer & Sims, 1994).

Many studies have been performed that evaluate the effectiveness of computer-aided instruction in the lab and the classroom (Anderson, 1987, 1993; Kulik & Kulik, 1987; Lee, 1992; Rosen & Petty, 1992). With its ability to combine computer-aided instruction with audiovisual media, interactive multimedia moves beyond traditional computer-aided instruction and has created new interest in educational technology (Hapeshi & Jones, 1992; Latchem,

Williamson, & Henderson-Lancett, 1993).¹ However, most studies in multimedia-based training have not been controlled and have focused on student satisfaction rather than on examining what students have learned (Reeves, 1992, 1993; for exceptions, see Cognition and Technology Group, 1993; Mayer & Sims, 1994). Some articles focus solely on development issues without consideration of evaluation (Barker, 1989; Hapeshi & Jones, 1992). Reeves (1993) claims that the design of multimedia has been driven by "habit, intuition, prejudice, guesswork or politics" (p. 79). For example, in a recent prescriptive article for multimedia development, Hapeshi and Jones (1992) describe why multimedia is so promising and what features the multimedia should possess, but they provide no experimental data. Given that some studies of computer-aided instruction have not found a benefit for the replacement of an instructor, the same scrutiny given to computer-aided instruction should be given to multimedia-based instruction (Welsh & Hull, 1991).

In addition to multimedia, alternatives to classroom teaching, such as video teletraining, have become increasingly popular (Dingus & Gillan, 1991; Miller, 1991). Research describing the benefits of multimedia, as compared with traditional or other forms of instruction, such as video teletraining, could benefit the educational technology community. Thus, the goal of this set of experiments was to evaluate the efficacy of a multimedia laboratory, as compared with traditional and teletraining, in producing increased learning and retention of both factual knowledge and skills in experimental methods.²

EXPERIMENT 1

In a course on learning, students need a laboratory to train rats in order to grasp the abstract principles that are taught. With the development of Sniffy the rat (Graham et al., 1994), this opportunity for hands-on learning can be provided for psychology students who study in a department without animal laboratories.

This experiment was designed to test whether a multimedia-based laboratory could increase learning on two different topics, as compared with a recitation-style

This work was partially funded by the Center for Teaching Excellence. The authors thank Evan Upchurch, Shane Melton, Krisela Rivera, and Erik Troberg for their help on various aspects of this project. Correspondence should be addressed to A. Y. Lee, Department of Psychology, New Mexico State University, Las Cruces, NM 88003 (e-mail: alee@nmsu.edu).

lecture. In the recitation, a small group of students listened to the same material presented by a graduate teaching assistant. Thus, the recitation resembled a traditional lecture, but was smaller in size.

Method

Design. A within-subjects design was used. For the first half of the semester, half the subjects used the operant conditioning software and half listened to a recitation-style lecture on the same topic. In the second half of the semester, the students who heard the recitation in the first half used software on eyewitness memory and those who used operant conditioning software listened to a recitation on eyewitness memory.

Subjects. Sixty-four undergraduate students at New Mexico State University taking an upper division course on learning were used. Five students did not complete the experiment.

Materials. Two software packages were used in this experiment. The first software package was Sniffy the rat, a simulated laboratory module designed to train subjects on the principles of operant conditioning (Graham et al., 1994). We developed our own accompanying background material and exercises for students to do. The students were asked to observe the natural behavior of the rat and make recordings on ethnogram sheets. They then magazine-trained the rat, shaped the rat to barpress, and, finally, varied the schedules of reinforcement. The students recorded observations and answers to questions on sheets of paper. No feedback was provided by the software except for the change in behavior observable in the rats.

The second software package, developed at NMSU, was designed to train students on experimental methodology, using an experiment on eyewitness memory as its basis (Loftus, Miller, & Burns, 1978). The program stepped the students through each part of designing materials and running an experiment. Buttons were used to allow students to obtain additional information or definitions of terms. If the students made mistakes in their selection of materials (ordering of critical items on a questionnaire, numbers of subject, design, and counterbalancing), they received feedback as to the correct response from the software. The subjects could not skip sections of the experimental setup, and the running of the experiment was animated. Results were automatically displayed and consisted of a randomly chosen table of data from a set of tables in the software. In this way, the students often received a set of data different from their neighbors' set.

For the recitation-style lectures, materials (lecture notes and overheads) covered the same concepts as those in the software pro-

grams. Variations in the materials between software and recitation may have resulted from the students' questions to the lecturer.³

To determine software usability, at the end of the semester, all students were given a questionnaire that asked them to assess various aspects of the eyewitness software.⁴ This questionnaire consisted of both evaluation and diagnostic questions concerning the software.

Procedure. The same basic procedure was used for the operant conditioning and eyewitness memory training. Before the labs were started, the subjects were given a pretest on the topic being trained. Then, over a 4-week period, the subjects were brought into the lab in groups for training (2 students per computer; 8 students listening to the recitation). After all students had completed the lab, a posttest on the topic was given. In addition, the subjects were given a questionnaire and cognitive ability tests.

Results and Discussion

Data collected. Three types of data were collected: pretest and posttest scores and usability questionnaire scores. The subjects' responses to pretests and posttests for both operant conditioning and eyewitness memory were scored for number correct.

Operant conditioning results. The subjects in both conditions (computer and recitation) improved from pretest to posttest [$t(57) = 6.4, p < .01$]. A mixed-model multiple regression was performed with test (pretest and posttest) as the within-subjects factor and condition (computer or recitation) as the between-subjects factor. (The regression equations for both operant conditioning and eyewitness memory are plotted in Figure 1.) A trend in the data indicates that for the recitation, the lecture helped the subjects scoring low on the pretest more than those scoring high on the pretest; however, for those using the computer, multimedia helped the subjects equally across the range of pretest scores [$F(1,57) = 3.22, p < .08$]. This result may have been due to the large number of conceptual questions on the pre- and posttests.

Eyewitness memory results. The subjects in both conditions (computer and recitation) improved from pretest to posttest [$t(58) = 7.8, p < .01$]. A mixed-model regression analysis was performed with test (pretest and

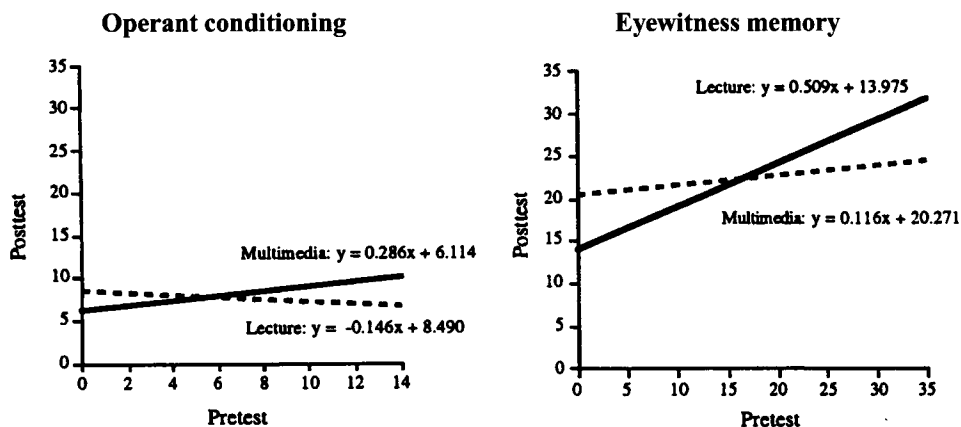


Figure 1. Pretest and posttest results for the operant conditioning and eyewitness memory programs for Experiment 1.

posttest) as the within-subjects factor and condition (computer or recitation) as the between-subjects factor. For those using the computer, the subjects who scored lower on the pretest improved more than did those who scored high on the pretest; however, for those in recitation, the subjects were helped equally across the range of pretest scores [pretest \times condition interaction, $F(1,58) = 4.08$, $p < .05$].

Usability results. Overall, the users rated the usability of the eyewitness software to be high. The median ratings on the two overall usability questions—ease of use and how well the users liked the system—were each 6 on a 7-point scale. Of the 25 users who completed questionnaires, 2 rated the ease of use to be below the neutral point (both rated it as 3), 3 gave a neutral rating, and 20 rated it to be above neutral (with 17 rating their liking to be a 6 or a 7) ($p < .0001$, sign test). A multiple regression analysis relating diagnostic questions concerning text, video, animation, controls, and navigation to overall ratings showed only ease of reading text to have a significant positive relation, accounting for 35% of the variance in the overall ratings. Accordingly, subsequent iterations of the design will focus on improving the text and replacing some text with graphics.

EXPERIMENT 2

Software is not the only medium used as a supplement to classroom teaching. Video telelearning has become a substitute for classroom learning for many students around the country (Dingus & Gillan, 1991). The video medium provides an inexpensive way to transmit information and allows students the flexibility to listen to a lecture wherever they are (e.g., on a navy ship, in a corporation's conference room, or at home; Miller, 1991; Tobagi, 1995). Video telelearning consists of components similar to those of multimedia-based training and does not require an instructor to be present. In addition to video telelearning, many instructors augment their class lectures with videos. For example, a lecture in introductory psychology on clinical interviews benefits from a video showing a practitioner conducting such an interview. With an increase in video telelearning and videos as supplements to lectures, a comparison between these two ways to augment classroom instruction—multimedia and video—can provide information about the effective use of both. Thus, Experiment 2 was designed to test whether a multimedia-based laboratory could increase learning on operant conditioning and eyewitness testimony, as compared with a video lecture on the same material.

Method

Subjects. Forty-two undergraduate students at NMSU taking an upper division course on experimental methodology were used. One student did not complete either part of the experiment. The students who completed at least one set of the materials (pretest, train, posttest) were included in the data even if they did not complete both topics.

Materials. The materials were similar to those used in Experiment 1, with the following changes. For the operant conditioning part of the labs, material on schedules of reinforcement was removed from both the operant conditioning lab and the operant conditioning recitation video. The videos consisted of the same material used for the recitations in Experiment 1. In addition, the same teaching assistant was used in the videos.

Both tests for operant conditioning were revised with more procedural questions, as well as a question on interpreting behavior graphs. In addition, a usability questionnaire similar to the eyewitness software questionnaire described for Experiment 1 was used. For the eyewitness testimony part of the labs, the software was changed to reduce the amount of text and integrate more effectively the video segment of a car accident into the body of the computer instruction.

Design and Procedure. The design and procedure were identical to those of Experiment 1, except that the eyewitness software was tested first. The video was shown in the same manner as were the recitations held in Experiment 1. The students could ask questions of the laboratory teaching assistant; however, the students did not have control of the VCR (e.g., could not replay sections or stop the video).

Results and Discussion

Data collected. Only data from the students who completed pretest and posttest for a lab (eyewitness memory or operant conditioning) were used. Three types of data were collected: pretest and posttest scores and usability questionnaire scores. The subjects' responses to pretests and posttests for both operant conditioning and eyewitness memory were scored for number correct. The mean scores for pretest and posttest for both operant conditioning and eyewitness memory are shown in Figure 2.

Operant conditioning results. The subjects in both conditions (computer and video) improved from pretest to posttest [$t(37) = 6.29$, $p < .01$]. A mixed-model multiple regression was performed with test (pretest and posttest) as the within-subjects factor and condition (computer or video) as the between-subjects factor. No difference was found between groups [$F(1,37) = 0.05$, $p = .96$]. This result may have been due to the large number of conceptual questions on the pre- and posttests. Thus, regardless of condition, the subjects improved from pretest to posttest. No interaction between condition and pretest was found.

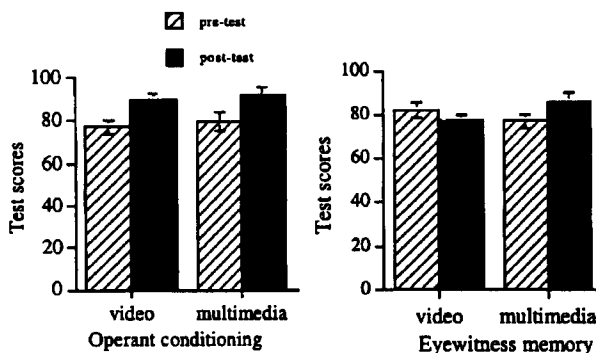


Figure 2. Pretest and posttest results for the operant conditioning and eyewitness memory programs for Experiment 2.

Eyewitness memory results. A mixed-model regression analysis was performed with test (pretest and posttest) as the within-subjects factor and condition (computer or recitation) as the between-subjects factor. The subjects who used the computer improved from pretest to posttest, whereas those who watched the video did not improve [mean difference for computer = +8.60; mean difference for video = -5.00; $F(1,37) = 2.09, p < .04$]. When condition is taken into account, then pretest predicts posttest performance [$F(1,37) = 1.99, p < .05$]; however, overall, the subjects did not improve between pretest and posttest [$t(37) = 0.75, p = .46$]. No interaction between condition and pretest was found.

Usability results. Overall, the users rated the usability of the Sniffy software to be high, with medians of 6 (on a 7-point scale) for both ease of use and preference questions (sign test to determine positivity of the ratings, both $ps < .01$). In contrast (and in contrast to Experiment 1), the subjects who rated the eyewitness program were not as uniformly positive, giving median ratings of 4 on both the ease of use and the preference questions. This difference between Sniffy and the eyewitness programs, as well as between Experiments 1 and 2 for eyewitness, may have been a function of the generally high knowledge of the subjects in Experiment 2.

Multiple regression analyses relating diagnostic questions to the overall questions—ease of use and preference—showed only significant relations between diagnostic questions and ease of use for the Sniffy program [$F(6,12) = 5.69, p < .01$]. Among the diagnostic questions, one feature of the computer program—the ability for the user to tell what actions to perform in order to get Sniffy to do something—was significantly related to ease of use ($t = 3.28, p < .01$) and accounted for 23% of the variance. Two features of the accompanying background material—the amount of information and the ability to skip information—were significantly related to ease of use ($ts = 2.86$ and 3.60 , respectively, both $ps < .05$) and together accounted for 37% of the variance.

GENERAL DISCUSSION

Measuring the effects of change in knowledge rather than just motivation or preference is better for evaluating educational efficacy.

For Experiment 1, the two multimedia-based training systems were compared with a recitation-style lecture on the same topics. The operant conditioning and eyewitness memory training produced very different results. In the operant training, the recitation provided substantial benefit to students who scored low on the pretest. In contrast, in the eyewitness memory training, the multimedia software provided the most benefit to students who scored low on the pretest. In other words, in some cases, replacing recitations with a multimedia lab may benefit students with low initial knowledge, but, in other cases, a recitation may be better.

For Experiment 2, the two multimedia-based training systems were compared with a video-style lecture on the same topics. Again, the operant conditioning and eye-

witness memory training produced different results. In the operant training, students improved whether they were given the multimedia program or the video. On the other hand, in the eyewitness memory training, students in the computer condition improved from pretest to posttest, whereas students in the video condition appeared to do worse.

Possible Explanations for Differences

The differences in results between the two experiments can be explained partially by the comparisons used and partially by the differences in the pool of subjects. In Experiment 1, the recitation-style lecture was given by a teaching assistant. In Experiment 2, a videotape of the teaching assistant was used. Although a teaching assistant was available to answer questions the students may have had, the interaction was not the same. In fact, questions allow an instructor to elaborate on materials in more depth and tailor the material to the class. Thus, a difference in materials between a live classroom and a computer program would necessarily exist. In the same way, with video students cannot interrupt the instructor when they may need more clarification. (Please note that these differences may also occur even when a teacher teaches the same materials to two different classes.) Thus, the challenge for a software developer is to anticipate possible places where students would need more clarity or even places where students may want to explore the topic beyond what the average student might need (to pass the class or an exam).

An additional issue is that, for Experiment 2, the high scores on the pretests for both operant conditioning and eyewitness memory indicate that there were no students with low initial knowledge. If students with low initial knowledge had been included, more similar results might have been obtained between Experiments 1 and 2; however, in the second semester, many students have had courses that cover similar topics (e.g., behavior modification, cognitive, memory, developmental methods). Thus, adaptive computer software that can take into account changes in students' knowledge may be beneficial for labs that supplement courses.

The differences in results between the operant conditioning and eyewitness memory training may have been caused by (1) software differences, such as amount of feedback for student actions, degree of interactivity, and goal-directedness of instruction, (2) test (pretest/posttest) differences, such as number of procedural versus number of conceptual questions, and (3) individual differences, such as intelligence or spatial abilities. However, preliminary analyses of the cognitive ability tests and intelligence scores suggest that these individual difference factors did not contribute to the present findings. Furthermore, in Experiment 2, the number of procedural and conceptual questions were balanced for the tests. Thus, the primary differences in results between the two topic areas may have been in the software. These results imply that each new piece of multimedia software may need its own evaluation in order to determine its effectiveness,

unless a set of empirically derived multimedia-based training guidelines are developed.

Although some discussion about multimedia guidelines can be found (Barker, 1989; Hapeshi & Jones, 1992), no multimedia guidelines have been extensively evaluated for whether they actually created systems that would improve learning in a manner similar to Anderson and colleagues' guidelines for intelligent tutoring systems (Anderson, 1987, 1993). Related work can be found in Mayer's (1989) studies for understanding how illustrations improve text comprehension and how mental models can be used to teach many different topics. His ideas indicate that learners' initial knowledge may affect how they use a system and learn from it. In support of a focus on the individual's knowledge base, Gay (1986) found that students with low conceptual knowledge learned more from computer-assisted video instruction when they were directed through the material and that students with high conceptual knowledge functioned well in both directed and undirected conditions. Students with high conceptual knowledge made better use of their time and also were able to focus on the relevant material for their own needs. Thus, further research should be performed in the area of developing multimedia guidelines. In the meantime, a developer could combine the ideas of Mayer and Anderson and focus on ways to create multimedia training software that adapts to individual students' previous knowledge level.

Conclusion

In summary, there are differential effects in how much people learned from multimedia (or other types of supplements to lectures) for people with different levels of initial knowledge. The operant conditioning software was less beneficial for students with less initial knowledge in the topic than was the eyewitness software. For training such students, multimedia training software may need to include more aid, such as more goal-directedness, feedback on activities, and connections with previous knowledge. Although students with high initial knowledge appear to benefit from some types of multimedia, they may not benefit from video supplements to courses. Finally, the effects of multimedia may be subtle. Therefore, evaluations of multimedia must be carefully performed in order to prevent prematurely accepting or rejecting a piece of software. In addition, the area of multimedia would benefit from a set of tested guidelines such as those created for intelligent tutoring systems.

REFERENCES

- ANDERSON, J. R. (1987). Methodologies for studying human knowledge. *Behavior & Brain Sciences*, *10*, 467-505.
- ANDERSON, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- BARKER, P. (1989). Multi-media CAL. In P. Barker (Ed.), *Multi-media computer assisted learning* (pp. 13-43). New York: Nichols.
- COGNITION AND TECHNOLOGY GROUP. (1993). Anchored instruction and situated cognition. *Educational Technology*, *33*, 52-70.
- DINGUS, T. A., & GILLAN, D. J. (1991). The thesis simulation: An approach for teaching research skills in a remote non-thesis program. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 505-507). Santa Monica, CA: Human Factors Society.
- GAY, G. (1986). Interaction of learning control and prior understanding in computer-assisted video instruction. *Journal of Educational Psychology*, *78*, 225-227.
- GRAHAM, J., ALLOWAY, T., & KRAMES, L. (1994). Sniffy, the virtual rat: Simulated operant conditioning. *Behavior Research Methods, Instruments, & Computers*, *26*, 134-141.
- HAPESHI, K., & JONES, D. (1992). Interactive multimedia for instruction: A cognitive analysis of the role of audition and vision. *International Journal of Human-Computer Interaction*, *4*, 79-99.
- KULIK, J. A., & KULIK, C. C. (1987). Review of recent literature on computer-based instruction. *Contemporary Educational Psychology*, *12*, 222-230.
- LATCHEM, C., WILLIAMSON, J., & HENDERSON-LANCETT, L. (1993). IMM: An overview. In C. Latchem, J. Williamson, & L. Henderson-Lancett (Eds.), *Interactive multimedia: Practice and promise* (pp. 19-38). Philadelphia, PA: Kogan Page.
- LEE, A. Y. (1992). Using tutoring systems to study learning: An application of HyperCard. *Behavior Research Methods, Instruments, & Computers*, *24*, 205-212.
- LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 19-31.
- MAYER, R. E. (1989). Models for understanding. *Review of Educational Research*, *59*, 43-64.
- MAYER, R. E., & SIMS, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*, 389-401.
- MILLER, D. (1991). Trim travel budgets with distance learning. *Training & Development*, *45*(9), 71.
- REEVES, T. C. (1992). Evaluating interactive multimedia. *Educational Technology*, *32*, 47-53.
- REEVES, T. C. (1993). Research support for interactive multimedia: Existing foundations and new directions. In C. Latchem, J. Williamson, & L. Henderson-Lancett (Eds.), *Interactive multimedia: Practice and promise* (pp. 79-96). Philadelphia, PA: Kogan Page.
- ROSEN, E. F., & PETTY, L. C. (1992). Computer-aided instruction in a physiological psychology course. *Behavior Research Methods, Instruments, & Computers*, *24*, 169-171.
- TOBAGI, F. A. (1995). Distance learning with digital video. *IEEE Multimedia*, *2*, 90-93.
- WELSH, J. A., & NULL, C. H. (1991). The effects of computer-based instruction on college students' comprehension of classic research. *Behavior Research Methods, Instruments, & Computers*, *23*, 301-305.

NOTES

1. The broadest definition of multimedia is used. Some researchers (e.g., Barker, 1989) use a very broad definition encompassing many activities, including computer, other equipment, and noncomputer activities, whereas others (e.g., Hapeshi & Jones, 1992) suggest that multimedia must contain an audio and visual mix. For our purposes, multimedia consists of simulation, audiovisual information, and computer and noncomputer activities.

2. We are not replacing instruction with multimedia but instead are augmenting lectures. Students received regular lectures on the topics in their classes, and the multimedia-based training provided hands-on experience outside of the classroom.

3. We did not control for student questions because we wanted the interactions between students and recitation instructor to be similar to a real recitation situation.

4. For Experiment 1, the focus of the usability evaluation was on the eyewitness software because we developed it. However, for Experiment 2, we asked the subjects to complete a usability evaluation for both the eyewitness and the operant conditioning software.