Behavior Research Methods, Instruments, & Computers
2000, 32 (2), 230-239

ARTICLES FROM THE SCiP CONFERENCE

Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses

RUTH H. MAKI, WILLIAM S. MAKI,
MICHELE PATTERSON, and P. DAVID WHITTAKER
Texas Tech University, Lubbock, Texas

We offered introductory psychology on the World-Wide Web (WWW) and evaluated the on-line format relative to the traditional lecture–test format, using a pretest–posttest nonequivalent control group design. Multiple sections of the introductory course were offered each semester; on-line and lecture sections were taught by the same instructor, the same textbook was used, and the same in-class examinations were taken. For on-line sections, mastery quizzes, interactive individual exercises, and weekly laboratory meetings replaced lectures. Increased content knowledge was greater for the students in the Web sections, as was in-class examination performance. Use of the WWW and computers for academic purposes increased more in the on-line sections, and the on-line students showed a greater decrease in computer anxiety. The students in the on-line sections expressed appreciation for course components and the convenience of the course, but the lecture sections received higher ratings on course evaluations than did the on-line sections. Learning and course satisfaction were dissociated in the two course formats.

Information technology appears to hold great promise for improving learning and increasing access to higher education via distance learning opportunities (see, e.g., Bork, 1997; Dede, 1996). However, despite a long history of computer-assisted instruction and distance education, there is little well-controlled empirical research regarding the effects of such technology. In two recent reviews, one on distance learning (Merisotis & Phipps, 1999) and one on hypermedia (Dillon & Gabbard, 1998), few papers remained after reports of interface design, opinion pieces, and reports of research lacking control groups were excluded. Thus, even while the potential of information technology continues to grow, our knowledge about its effects on learning remains scant. Such knowledge, gained from sound designs, is needed in order to deploy instructional technology most effectively.

We developed an introductory psychology course, using the World-Wide Web (WWW) as an alternative to a lecture format.[1] This alternative format addressed several problems associated with large introductory courses (see, e.g., Wilson, 1996), including poor attendance (averaging about 60%, according to Forsyth & Archer, 1997), and disinterest and inappropriate behavior (talking, sleeping, reading) on the part of students who do attend class. Furthermore, the practice of infrequent testing generates sporadic bursts of study activity prior to tests (see Wilson, 1996), and research has shown massed study to be less effective than study distributed across time (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993).

In our view, these problems are symptomatic of the *teaching* paradigm, which is the philosophical basis for the one-to-many, lecture–test method (Barr & Tagg, 1995). There is no pedagogical rationale for teaching in the lecture format. The reasons are historical, traditional, and economic. For example, textbooks, films, and lectures used to be the only media available to college and university lecturers. With the advent of other types of media and the explosive growth in information technology, we have the opportunity to effect a shift to a *learning* paradigm focusing on students' learning experiences.

On the basis of the learning paradigm, we designed an alternative format for teaching introductory psychology that takes advantage of the technology presently available. Much of the coursework was done at the students' convenience, using personal computers, located either on or off campus. We offered this on-line version of the introductory psychology course during the fall and spring semesters, beginning in August 1998. Students studied a standard textbook (Kalat, 1996). In lieu of lectures, however, our course included computerized presentation of

Portions of this research were presented at the meetings of the American Psychological Association, August, 1999, Boston, and at the Society for Computers in Psychology, November, 1999, Los Angeles. The project was supported by Grant NSF DUE-9752349 from the National Science Foundation to R.H.M. and W.S.M. ("Using Technology to Transform Instruction and Improve Learning in Introductory Psychology"). Correspondence concerning this article should be addressed to R. H. Maki, Department of Psychology, Texas Tech University, Lubbock, TX 79409-2051 (e-mail: ruth.maki@ttu.edu).

course materials on the WWW, "laboratory" exercises that included computerized demonstrations of psychological phenomena and searching the Web for information, interactive mastery quizzes with feedback, and e-mail contact between students, the instructor, and the teaching assistants.

We selected the components of our course that were expected, on the basis of published literature, to either increase student learning or make large classes more personalized. Forsyth and Archer (1997) reported the use of a technologically enhanced classroom in an introductory psychology course. The course included on-line lecture notes, on-line chapter outlines, a computerized study guide, e-mail connection to the instructor, and multimedia learning exercises. Students rated the computer-based components of the course quite favorably. They indicated that the technological aspects of the course were a positive learning experience. Although the distribution of grades in the technologically enhanced class was similar to that for lecture courses from previous years, heavier users of the technology earned higher grades. In addition, weaker students (as judged by self-reported high school grades) who used the technology more scored higher on examinations than did weaker students who used it less. However, these relationships may show more about motivation than about learning.

Worthington, Welsh, Archer, Mindes, and Forsyth (1996) used computer-assisted instruction (CAI) to add simulations, demonstrations of classic experiments, and tutorials to a lecture class. Another lecture class did not receive the CAI enhancements. Those students who participated in the CAI section increased scores from a pretest to a posttest more than did the students in the no-CAI class. This was especially true for test questions related to the computerized exercises. A majority of the students (65%) said that they preferred the CAI-enhanced course over a standard lecture course.

In contrast to the above studies in which CAI supplemented lectures, other studies have compared largely on-line courses with traditional lecture courses. In almost all direct comparisons to date, measures of student learning and satisfaction in computer-assisted learning courses have been equal to or higher than those obtained from traditional courses (Hiltz, 1993; Schutte, 1996; cf. Magnuson-Martinson, 1995). For example, Schutte compared on-line versus face-to-face versions of a statistics course and found that the on-line students performed better on a common final examination. Spooner, Jordan, Algozzine, and Spooner (1999) compared end-of-course evaluations for a special education course that was offered on campus and off campus that used various types of electronic media. They found that course and instructor ratings were similar across types of course.

Pear and Novak (1996) conducted two psychology courses using a computer-aided personalized system of instruction. Students only met for the first 2 weeks, and then the entire course was delivered on computers. Students studied the textbook, but they took short-answer quizzes with randomly selected questions on the computer as often as necessary until they met a mastery criterion. The features of the course that the students liked best were its convenience and not having to attend classes. Features they liked least were the heavy weighting of the conventional examinations, the lack of interaction between themselves, other students, and the professor, the absence of classes, and the specific questions on the examinations. About a third of the students expressed some dissatisfaction with the course, and the majority of those commented on problems in using the computers as the source of dissatisfaction. The other two thirds were generally satisfied with the course, and about three quarters of the students indicated that they would take such a computerized course again.

Our evaluation compared a class that was mostly taught on the WWW with a comparable lecture class. Students in both courses read the same textbook, took the same tests, and had the same instructors. For most of our measures, we used a pretest–posttest, nonequivalent control group (quasi-experimental) design. We measured learning, computer use and computer anxiety, and satisfaction.

METHOD

Design

In our quasi-experimental design, on-line sections of a general psychology class were the experimental groups, and lecture sections of the same course served as nonequivalent control groups. During each of the fall and spring semesters, 1998–1999, we offered four on-line sections of general psychology that had a maximum enrollment of 25 students, and we offered two lecture sections that had a maximum enrollment of 50 students. Measures of learning, attitudes toward computers, and computer use were collected at the beginning and the end of the course.

Graduate students in the doctoral program at Texas Tech University served as instructors for the lecture sections and as teaching assistants (TAs) for the on-line sections. One TA had 2 years of prior teaching experience, and one had 3 years. TAs were selected because they had been very successful as classroom teachers. All the students studied course material from a textbook (Kalat, 1996) and took the same three midterm examinations and a common final exam. Each TA was responsible for one lecture section and for the laboratory portion of two on-line sections each semester. We varied the activities in the laboratory meetings of the on-line course. During the fall semester, two of the sections (one for each TA) were assigned as demonstration sections. Those students participated in demonstrations and experiments that usually resulted in the pooling of data. This was followed by a class discussion, and groups of students wrote reports. The other two sections (one for each TA) were formed into study groups. Members of the study groups wrote questions about the course material, shared the questions with group members, and they spent their class time discussing the questions. During the spring semester, two sections were again assigned to be demonstration groups, but the study groups were replaced with review groups. These students were required to send two questions about course material to the TA via e-mail. The TA then presented all of the questions to the class during the weekly meeting and discussed topics that were most often included in the students' questions.

Participants

The participants were enrolled in general psychology at Texas Tech University. During each of the fall and spring semesters, there

were four on-line sections and two lecture sections. Because of university policy, the printed schedules contained asterisks beside the on-line sections, indicating that they were largely taught using the WWW. Despite this, many students expressed surprise at the nature of the course. The students in all the sections were informed that we would be collecting some data to evaluate the course, and they signed a release agreeing that their data could be used.

During the fall semester, 1998, the initial enrollment in the on-line sections was 70, and the final enrollment was 59. Fall enrollment in the lecture sections was originally 97, with a final enrollment of 93. During the spring semester, 1999, the initial enrollment was 81 students in the on-line sections, with a final enrollment of 71. The lecture sections began with 85 students and ended with 82. The withdrawal rate from the on-line sections (13.9%) was significantly greater than the withdrawal rate from the lecture sections [3.8%; $\chi^2(1) = 10.87$, $p < .05$][2]. However, the class standings of those students who withdrew were similar for the two formats: About 55% of the withdrawals were freshmen, 30% were sophomores, and 15% were juniors and seniors. The composition of the sections varied. The on-line sections included 3.4% psychology majors, and the lecture sections included 11.0% psychology majors, and this difference was significant [$\chi^2(1) = 5.55$].

Procedure

The students in the lecture sections met three times a week with their graduate student instructors, who taught the course in the same way as they had in previous years with a lecture format. In the on-line sections, the students met during the regular class times during the first 2 weeks. These meetings were used to describe the course, to teach the use of a browser, to show students the course structure, and to teach e-mail use. Precourse evaluation data were collected from both the on-line and the lecture sections during the 2nd week. Postcourse evaluation data were collected from all the sections during the last week of classes. After the orientation sessions, the on-line groups met on Wednesday of each week to participate in the demonstration, study group, or review activities or to take a scheduled exam. On-line work involving computerized demonstrations and experiments and interactive quizzes replaced the other two scheduled class sessions.

Computer assignments. Textbook reading assignments were posted to the Web, linked to an interactive course syllabus, along with chapter outlines, the textbook's glossary, and links to related Web sites. Frequently asked questions (FAQs) were linked to the outline for each chapter. The instructor posted FAQs and answers related to material that prior students had found difficult. During each semester, review questions were embedded in either the chapter outlines or the FAQs. The purpose of these questions was to reward the students for using either the chapter outlines or the FAQs. The students received credit for correctly answering these questions. The results of this manipulation are described in a companion paper (Maki & Maki, 2000).

Students in the on-line course used computers to work on two types of course-related assignments in lieu of lectures: (1) conducting computerized demonstrations and searching the Web for information related to the course, and (2) taking mastery quizzes on the Web.

Each week, the students participated in one interactive assignment. Computerized demonstrations included brief experiments on such topics as visual illusions and measurement of memory. For perception, for example, the students measured the size of the Poggendorff illusion, manipulating the size and color of the bar separating two diagonal lines. They adjusted one of the diagonals so that it appeared to be a continuation of a diagonal on the other side of the bar. After collecting data, the students answered questions about the exercise, and these, along with the individual's data, were automatically e-mailed to the course instructor. Web searches replaced computerized experiments for some chapters, because material in those chapters did not lend itself to experimental demonstrations. For these chapters, the students were assigned to find information on the Web and answer questions about it. For example, for the introductory chapter, the students were asked to visit the Web sites of the psychology departments of two universities and to answer questions about the specializations of the faculty in those departments.

Each week, the students were also required to pass (at 80% or better) two mastery quizzes. Quizzes contained test questions drawn from the same pool as actual test questions. The test item pool furnished by the publisher contained about 100–150 test questions per chapter. This was split in half, by odd and even questions. One half made up the pool of quiz items, and the other half served as exam items. Using the same pool for mastery quizzes and actual tests ensured that the students received questions that were very similar to the test questions. Furthermore, the students were instructed to study each chapter completely before taking the mastery quiz. A package of Perl scripts (Schwartz, 1998; Wall, Chistiansen, & Schwartz, 1998) was written to administer mastery quizzes. Each quiz consisted of 15 randomly sampled questions for each chapter. The students saw the multiple-choice questions, made their choices, and then received feedback, including reasons why the alternatives were correct or incorrect and the textbook pages and sections of the textbook containing relevant information. The students were encouraged to make notes on questions that they answered incorrectly and to note the pages and topics that needed more study. Although the students were required to pass two quizzes each week, they could take as many quizzes per chapter as they wished. A database was kept for each student so that he or she would receive new questions on each retake. Once questions from a pool had been used, the students received questions they had answered wrongly on previous mastery quizzes.

Credit for completing the demonstrations and quizzes was automatically entered into a database, to keep student records. At any time, the students could request a course progress report that contained all of their credit information.

Measures

The students in the lecture and the on-line sections answered a precourse and a postcourse questionnaire. These consisted of a computer anxiety scale, a computer use scale, personality scales, and 3–4 questions related to each textbook chapter taken from the practice test for the psychology GRE (Educational Testing Service, 1994). In addition, the students in lecture and on-line sections took the same four midterm and final examinations. Finally, all the students answered questions about their satisfaction with the section of general psychology that they took.

Content questions. The psychology GRE practice test booklet (Educational Testing Service, 1994) consists of 320 multiple-choice questions. Practice GRE questions were selected because they were publicly disclosed questions that had appeared on previous GRE tests. Thus, the questions had been judged to be both reliable and valid, using the Educational Testing Service's standards for GRE tests. We selected 40 questions to tap into material that is covered in general psychology. The students in the on-line and lecture sections answered these questions before and after the course.

Examinations. The students in the lecture and on-line sections took four unit examinations, covering 3–4 chapters each. These examinations consisted of multiple-choice questions taken from those questions from the test bank that were not used in mastery quizzes for the on-line students. Questions that covered material taught by the TAs in their lecture courses were nominated by the TAs. The students in all the sections also took a comprehensive multiple-choice final exam.

Measures of computer use. Because computer usage plays a major role in the on-line course, we used two computer-related scales. One was a 16-item computer anxiety scale (Cohen &

Table 1
Mean Percentages Correct and Standard Errors of the Mean (*SE*s) on Content Questions and Course Examinations for On-Line and Lecture Sections

	On-Line Sections		Lecture Sections	
Measure	*M*	*SE*	*M*	*SE*
GRE Questions				
Pretest	25.51	0.88	26.78	0.70
Posttest	35.63	1.35	32.18	0.83
Class Examinations				
Exam 1	73.22	1.32	71.39	1.03
Exam 2	71.86	1.34	69.59	1.02
Exam 3	76.49	1.27	72.01	1.01
Exam 4	67.53	1.19	64.69	1.10
Final Exam	73.37	1.25	75.50	0.94

Waugh, 1989), which included such items as "I feel anxious whenever I use computers" and "I wish that computers were not as important as they are." These items were rated on a 5-point scale, varying from *strongly disagree* to *strongly agree*. Some items were positive with respect to computers and those were reverse scored. The second scale was a computer use scale tapping four dimensions of people's perceptions of computers and their use (Panero, Lane, & Napier, 1997). The Computer Enthusiasm scale included such questions as "I shop for computer hardware or software by going to stores or looking at catalogs" and "I spend time configuring the computer to look and act as I want it to." The Efficiency in Work scale included such questions as "I use a computer to save time on work that would take me longer otherwise." The Entertainment scale included such items as "I play games on a computer." Finally, the Communication scale included such questions as "I use a computer to keep in touch with friends and family who are far away." We also asked questions about frequency of use of computers in college classes and use of e-mail and the WWW. Ratings were made on a 5-point scale, varying from *never* to *very often*.

Student satisfaction. We asked additional questions about the students' satisfaction and their perceptions of their psychology course. These included the amount of time spent on the introductory psychology course, the likelihood of taking more psychology courses, how interesting the course was, whether the student would recommend the class to a friend, and whether he or she would sign up for the same section again. In addition, the students in the on-line sections were asked questions specifically related to the components of the course on a midterm course evaluation. They were asked both how much they used various parts of the course, how useful they were, and whether various components should be included in a redesigned course. The students were also asked to write what they liked most and disliked most about the course.

RESULTS

Each analysis was based on all the participants who had complete data on the measures used in the analysis. The number of participants varied, depending on whether we needed both pretest and posttest data and whether course evaluations had been turned in. For each analysis, the number of observations used from the on-line and the lecture methods is noted.

Mastery of Course Content

Both the in-class midterm examinations and the post-course measure of knowledge about psychology showed a small advantage in learning for students in the on-line sections, as compared with the lecture sections. Although the on-line and lecture sections started out about the same in knowledge about psychology, the on-line sections were able to answer more questions correctly at the end of the course, suggesting that the on-line students learned more content during the semester.

Scores on the GRE content questions at pretest and posttest for the on-line and the lecture sections are shown in Table 1. These data were based on 97 students from the on-line sections and 121 students from the lecture sections. A mixed design analysis of variance (ANOVA) was conducted, with pretest–posttest as the repeated measure and method of instruction, instructor, classification in college (freshman vs. all other), and semester as between-subjects variables. The students improved their scores from before to after the course [$F(1,202) = 79.89$, $MS_e = 59.24$], but more important, the on-line sections improved more than the lecture sections [$F(1,202) = 9.25$, $MS_e = 59.24$, for the interaction]. This differential change did not interact with instructor, semester, or classification [$Fs(1,202) \leq 2.12$, $MS_e = 59.24$]. At the beginning of the course, the difference between the lecture and the on-line sections was not significant ($F < 1$), but the on-line sections scored higher at the end of the course than did the lecture sections [$F(1,202) = 5.83$, $MS_e = 120.43$].

During their weekly meetings, the on-line sections participated in demonstrations and experiments (the demonstration group) or they participated in study groups during the fall semester and review sessions during the spring semester (the study/review group). The demonstration group improved their content scores by 9.25%, and the study/review group improved their scores by 11.32%. This difference for change scores was not significant [$F(1,90) = 1.08$, $MS_e = 74.21$, for group × time interaction].

Scores on four unit examinations are also shown in Table 1. These were analyzed in a mixed-design ANOVA, with exam number as a repeated measure and method of instruction, instructor, classification in college, and semester as between-subjects variables. This analysis was based on 116 students in on-line sections and 138 students in lecture sections for whom we had data on all variables. Overall, examinations differed in difficulty [$F(3,708) = 40.98$, $MS_e = 59.11$], but more important, as can be seen in Table 1, the students in the on-line sections performed better than students in the lecture sections [$Ms = 72.28\%$ vs. 69.42%; $F(1,236) = 5.55$, $MS_e = 470.16$]. This effect did not interact with other variables [largest $F(1,236) = 2.70$, $MS_e = 470.16$, for method × instructor × classification].

Overall, the demonstration and study/review groups scored similarly on the examinations [$Ms = 71.47\%$ and 73.13%, respectively; $F(1,100) = 1.58$, $MS_e = 540.59$]. However, there was an interaction among semester, group, and classification [$F(1,100) = 6.75$, $MS_e = 540.59$]. This interaction occurred because nonfreshmen (sophomores and above) in the review group scored significantly better than similar students in the demonstration group [$Ms = 80.72\%$ vs. 71.95%; $F(1,33) = 5.21$, $MS_e = 585.94$], but

Table 2
Mean Pretest and Posttest Scores and Standard Errors of the Mean (*SE*s) on Computer Anxiety, the Computer Use Scale (CUS), and Reported Frequency of Computer Use

	On-Line Sections				Lecture Sections			
	Pre		Post		Pre		Post	
Measure	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Computer Anxiety	1.97	0.05	1.80	0.06	2.30	0.04	2.16	0.06
Freshmen	2.20	0.09	1.83	0.08	2.22	0.08	2.08	0.07
Sophomore and above	1.79	0.08	1.78	0.09	2.45	0.11	2.33	0.10
CUS–Communication	3.22	0.12	3.48	0.12	2.76	0.11	3.00	0.11
CUS–Entertainment	3.07	0.08	3.25	0.09	2.69	0.08	2.77	0.08
CUS–Efficiency	3.96	0.06	4.08	0.05	3.67	0.06	3.69	0.06
CUS–Enthusiasm	2.32	0.09	2.51	0.09	1.87	0.07	1.96	0.07
Fall semester	2.18	0.14	2.50	0.14	1.86	0.10	1.95	0.09
Spring semester	2.46	0.13	2.51	0.13	1.88	0.10	1.97	0.10
E-mail frequency	4.12	0.12	4.43	0.09	3.40	0.12	3.95	0.11
WWW frequency	3.90	0.11	4.26	0.09	3.39	0.09	3.60	0.09
Computers in class	3.42	0.12	4.21	0.09	3.07	0.11	3.48	0.09

only in the spring, when the review procedure was used. This difference was not significant in the other conditions [$Fs(1,17) = 2.75$, $MS_e = 515.37$].

Scores on the cumulative final examination are also shown in Table 1. There was no significant effect of method of instruction ($F < 1$), and method did not interact with other variables [largest $F(1,236) = 2.80$, $MS_e = 144.28$, for the method × semester × instructor interaction]. Among the students in the on-line sections, there was a trend for the students in the study/review group to do better on the final exam than the students in the demonstration group [76.33% vs. 71.71%; $F(1,100) = 3.02$, $MS_e = 164.43$, $p < .10$]. As with midterm examinations, there was an interaction among semester, group, and classification [$F(1,100) = 4.01$, $MS_e = 164.43$]. Again, nonfreshmen in the review group in the spring semester scored higher on the cumulative final exam than did similar students in the demonstration group [82.28% vs. 72.83%; $F(1,33) = 5.53$, $MS_e = 145.62$]. Although there was a similar trend for freshmen in the spring and for all the students in the fall, none of the other comparisons was significant [$Fs(1,17) = 3.42$, $MS_e = 187.64$].

Use of and Attitudes Toward Computers

Each measure was collected at the beginning and the end of the semester. The analyses were based on 97 students in the on-line sections and 121 students in the lecture sections. Data were analyzed in mixed design ANOVAs, with time of measure as a repeated measure and method of instruction, semester, college classification, and instructor as between-subjects variables.

The average scores on the computer anxiety scale for the on-line and lecture groups at the beginning and at the end of their courses are shown in Table 2. Overall, the students showed less anxiety at the end of the semester than at the beginning [$F(1,202) = 18.88$, $MS_e = 0.12$]. Furthermore, the on-line group was less anxious than the lecture group [$F(1,202) = 22.25$, $MS_e = 0.69$], and there was a tendency for these two variables to interact with classification in college [$F(1,202) = 3.67$, $MS_e = 0.012$, $p < .06$]. As can be seen in Table 2, freshmen in the on-line sections dropped more, with time, on the anxiety scale than did freshmen in the lecture sections. For the freshmen, the interaction of time and method was significant [$F(1,114) = 5.95$, $MS_e = 0.012$]. The freshmen in the on-line and lecture sections did not differ in computer anxiety at the beginning of the course ($F < 1$), but the on-line students were less anxious than the lecture students at the end of the course [$F(1,114) = 6.63$, $MS_e = 0.31$]. For the more advanced students, neither the effect of time nor the interaction between time and method was significant ($Fs < 1$), but the students in the on-line sections were less computer anxious than the students in the lecture sections [$F(1,88) = 23.63$, $MS_e = 0.66$].

The Computer Use Scale evaluates how people use computers. It consists of four subscales—using computers for entertainment and for communication, belief that computers increase work efficiency, and computer enthusiasm. Means on each of these scales (5 being highest and 1 being lowest) are shown in Table 2. For entertainment, communication, and beliefs about efficiency, the on-line sections were significantly higher than the lecture sections [$Fs(1,202) \leq 22.79$, $MS_e = 0.60$]. However, this difference did not interact with time of measurement [largest $F(1,202) = 2.58$, $MS_e = 0.24$, for entertainment].

The students became more enthusiastic about using computers across the semester [$F(1,202) = 20.98$, $MS_e = 0.14$], and the students in the on-line sections were more enthusiastic than the students in the lecture sections [$F(1,202) = 25.41$, $MS_e = 2.19$]. However, time and method of instruction interacted with semester [$F(1,202) = 6.04$, $MS_e = 0.014$]. As can be seen in Table 2, the students in the on-line sections increased their enthusiasm for computers across the fall semester more than did the students in the lecture sections [$F(1,92) = 6.81$, $MS_e = 0.14$]. In the fall, the on-line and the lecture groups did not differ

Table 3
Mean Course Evaluation Ratings and Standard Errors of the Mean (*SE*s) for On-Line and Lecture Sections

Question	On-line Sections		Lecture Sections	
	M	*SE*	*M*	*SE*
How interesting is psychology?*	4.01	0.09	4.42	0.072
Recommend section to a friend†	3.45	0.14	4.36	0.08
Sign up for section again†	3.20	0.16	4.35	0.09
Had to work harder†	3.45	0.11	3.06	0.16

*Alternatives on this scale varied from 1, *not at all interesting*, to 5, *very interesting*. †These scales varied from 1, *strongly disagree*, to 5, *strongly agree*, so higher ratings indicate more agreement with the statements.

significantly at the beginning of the course [$F(1,110) = 1.79$, $MS_e = 0.82$], but the students in the on-line sections were more enthusiastic about computer use than were those in the lecture sections at the end of the course [$F(1,110) = 8.96$, $MS_e = 0.76$]. In the spring semester, time and method of instruction did not interact ($F < 1$), but the students in the on-line sections were more enthusiastic than the students in the lecture sections [$F(1,92) = 15.16$, $MS_e = 1.09$].

Computer usage questions asked how much students used e-mail, how much they had used computers in their college courses, and how often they had used the WWW to find information. Mean ratings on a 5-point scale (with 5 being *most frequent use*) are shown in Table 2. Each of these measures increased with time [smallest $F(1,202) = 17.81$, $MS_e = 0.50$, for use of the WWW]. In addition, the students in the on-line sections reported more use than did the students in the lecture sections [smallest $F(1,202) = 17.49$, $MS_e = 1.41$, for using computers in classes]. Reported use of the WWW and use of computers in classes increased more in the on-line sections than in the lecture sections [$F(1,202) = 4.19$, $MS_e = 0.050$, and $F(1,202) = 9.02$, $MS_e = 0.70$, respectively, for the interactions]. The students in the on-line sections reported more use of the WWW than did the students in the lecture sections both at the beginning [$F(1,202) = 9.27$, $MS_e = 1.12$] and at the end [$F(1,202) = 27.84$, $MS_e = 0.94$] of the course, but the change was greater during the course for the on-line sections. At the beginning of the course, the students in the on-line and the lecture sections did not differ significantly in their use of computers in classes [$F(1,202) = 2.60$, $MS_e = 1.15$], but the on-line students reported more class use of computers at the end of the course [$F(1,202) = 29.29$, $MS_e = 0.95$].

The students in the on-line sections decreased their anxiety more and increased their enthusiasm and use of computers more than did the students in the lecture sections. However, these effects were strongest early in the students' college experiences, either when they were freshmen (for anxiety) or during the fall but not the spring semesters (for enthusiasm). The on-line course may be especially effective for increasing appreciation for and use of technology for less experienced students. It appears that students with more experience selected the on-line course if they were already technologically sophisticated, in comparison with the students in the lecture sections.

Student Satisfaction

Overall course evaluations. Tables 3 and 4 show how the students evaluated the on-line and the lecture classes. The alternatives for most questions were on continuous scales (e.g., level of agreement), so means were calculated and questions were analyzed by using between-subjects ANOVAs, with method of instruction, semester, and instructor as variables. We collected these data anonymously during the fall semester, so we could not include college classification in the analysis.[3]

Students perceived psychology as more interesting when they took the lecture course than when they took the on-line course [$F(1,239) = 11.28$, $MS_e = 0.81$]; they would be more likely to recommend their specific section of the course to a friend [$F(1,239) = 28.44$, $MS_e = 1.42$], and they would be more likely to sign up for their section again [$F(1,239) = 38.34$, $MS_e = 1.74$]. *Interest* and *recommend* interacted with instructor [$F(1,239) = 5.71$, $MS_e = 0.81$, for the *interest* variable, and $F(1,239) = 14.47$, $MS_e = 1.42$, for the *recommend* variable]. For *interest*, one instructor produced a large and significant difference between the on-line and the lecture sections [Ms = 3.9 vs. 4.59; $F(1,122) = 21.76$, $MS_e = 0.66$], but the other instructor produced a smaller, nonsignificant difference (Ms = 4.19 vs. 4.30; $F < 1$). A similar pattern appeared for *recommend*, with one instructor producing a large difference [Ms = 3.20 vs. 4.69; $F(1,122) = 47.01$, $MS_e = 1.35$], and the other instructor producing a smaller, nonsignificant difference [Ms = 3.86 vs. 4.11; $F(1,117) = 1.04$, $MS_e = 1.49$]. Whether students would recommend their section to a friend also interacted with semester [$F(1,239) = 14.47$, $MS_e = 1.42$]. The difference between the lecture and the on-line sections was greater in the fall [Ms = 3.15 vs. 4.52;

Table 4
Percentage of Students Giving Various Responses in On-Line and Lecture Sections

Measure	Section	
	On-Line	Lecture
Amount of time per week on psychology course		
<2 h	17	20
2–4 h	49	40
4–6 h	26	33
6–8 h	7	6
>8 h	1	1
Likely to take more psychology courses		
Not at all likely	52	28
Probably take one course	30	35
Probably two or more courses	8	10
Surely take one course	3	9
Surely two or more courses	7	19

Table 5
Mean Perceived Usefulness* and Standard Errors of the Mean (*SE*s) for On-Line Course Components at Midterm by Demonstration and Study/Review Groups

	Demonstration Group		Study/Review Group	
Component	*M*	*SE*	*M*	*SE*
Quizzes	4.66	0.08	4.65	0.09
Chapter outlines	4.00	0.11	3.94	0.13
Frequently asked questions	3.82	0.11	3.96	0.13
Computer assignments†	4.10	0.11	3.65	0.13
Class meetings	3.54	0.13	3.88	0.14

*Each question was "How useful are _____ for learning course material?" with alternatives of 5, *very useful*, 4, *moderately useful*, 3, *slightly useful*, 2, *not at all useful*, and 1, *they interfered with learning*. †Demonstration group significantly higher than study/review group.

$F(1,28) = 28.52$, $MS_e = 1.29$] than in the spring [*M*s = 3.72 vs. 4.16; $F(1,116) = 4.96$, $MS_e = 1.46$], although both differences were significant.

Sign-up again produced a three-way interaction of method with instructor and semester [$F(1,239) = 6.18$, $MS_e = 1.74$]. One instructor produced higher ratings for the lecture course than for the on-line course in both the fall [*M*s = 4.91 vs. 2.37; $F(1,68) = 82.82$, $MS_e = 1.37$] and the spring [*M*s = 4.39 vs. 3.48; $F(1,54) = 5.48$, $MS_e = 2.03$]. The other instructor did not produce a significant difference between the lecture and the on-line sections in either the fall (*M*s = 3.98 vs. 3.59, $F < 1$) or the spring [*M*s = 4.20 vs. 3.72; $F(1,62) = 2.03$, $MS_e = 1.76$]. There were no significant effects of on-line group (demonstration vs. study/review) for any of the satisfaction measures (*F*s < 1).

In the fall, the on-line students were asked whether they agreed that they "had to work harder" in this course than in other courses; in the spring, both the on-line and the lecture students were asked this question. Stronger agreement by the on-line students than by the lecture students was marginally significant [$F(1,167) = 3.74$, $MS_e = 1.47$, $p < .06$].

The alternatives for the questions in Table 4 were not on a continuous scale, so they were analyzed as a function of instructional method with chi squares. Overall, the students spent about 4 h a week (including class time) on their psychology course, and this did not differ between the instructional methods [$\chi^2(4) = 2.02$]. As can be seen in Table 4, the students in the lecture course intended to take more psychology courses than did the students in the on-line course [$\chi^2(4) = 19.52$]. This may have been because there were more declared psychology majors in the lecture sections, or it may be that lectures increase interest in psychology and that leads to the intention to enroll in more courses.

Evaluation of on-line course components. A midterm evaluation questionnaire was answered by the students in the on-line sections during the 8th week of the semester. The students were asked how often they had used various features of the course. The modal response for use of the course syllabus with assignment dates was more than once a week. The modal response for "What's New" (the announcement page), the chapter outlines, and the FAQs was about once a week. The links to related sites on the WWW were never used by 46% of the students. We also asked the students to rate how useful the components of the course were for learning the course material. Along with a rating scale varying from *very useful* to *they interfere with learning*, we included "don't know—never use them" as an alternative. About 10% of the students selected that alternative for chapter outlines, but only 1 student selected it for the other course components. Mean ratings (excluding the "never use" responses) can be seen in Table 5. The features of the on-line course varied in perceived utility, with mastery quizzes considered to be very useful and other components rated as slightly to moderately useful.

We analyzed ratings for each component in 2 × 2 × 2 ANOVAs, with class activity (demonstration vs. study/review), instructor, and semester as between-subjects variables. The only significant effect of activity was for the computer assignments; the students in the demonstration group found them more useful than did the students in the study/review group [$F(1,104) = 8.81$, $MS_e = 0.89$], probably because many of the weekly class sessions elaborated on the computer exercises. No effects of activity interacted with semester or instructor.

Several questions about the on-line course were asked on the posttest questionnaire. These are shown in Table 6. Means above three show that the students were generally satisfied with specific aspects of the on-line course, even though they tended to give lower overall ratings than did those in the lecture courses. They felt that communication with the professors was better than in lecture courses. The students liked the convenience of the on-line format, and they agreed strongly that they liked the flexibility. They were neutral to slightly positive about whether the on-line course was a better learning experience than lecture courses, that it increased the efficiency of their education, and that use of the WWW increased the quality of their education. They were positive about learning on their own and coming to class only once a week. The students did not generally agree that more lectures were

Table 6
Mean Ratings and Standard Errors of the Mean (*SE*s) for On-Line Course by Demonstration and Study/Review Groups

Question*	Demonstration		Study/Review	
	M	*SE*	*M*	*SE*
E-mail improved access to professors	3.59	0.14	3.63	0.15
On-line class is more convenient than lecture course	3.76	0.18	3.50	0.19
I liked the flexibility of the on-line course	4.45	0.12	4.04	0.22
On-line format increased the efficiency of my education (I learned more in less time than in other classes)	3.23	0.17	2.88	0.18
This course was a better learning experience than lecture courses	3.29	0.16	2.77	0.17
Using the World-Wide Web increased the quality of my education	3.37	0.14	3.10	0.16
I liked having to learn on my own	3.91	0.21	3.70	0.20
I liked coming to class only once a week	4.39	0.17	4.26	0.22
I think more lectures are needed in the course	2.81	0.22	2.93	0.27
Would choose another on-line course†	3.63	0.18	2.98	0.20

*These scales varied from 1, *strongly disagree*, to 5, *strongly agree*, so higher ratings indicate more agreement with the statement. †Demonstration group significantly higher than study/review group.

needed. The students in the demonstration group were more likely than the students in the study/review group to choose another on-line course [$F(1,102) = 4.09$, $MS_e = 1.84$]. In the demonstration group, 62% of the students strongly agreed or agreed that they would select another on-line course; only 40% of the students in the study/review groups agreed or strongly agreed with this statement. There was also an effect of semester for this item. The students were more positive in the spring semester than in the fall semester [Ms = 3.74 vs. 2.85; $F(1,102) = 15.78$, $MS_e = 1.84$], but semester did not interact with type of activity ($F < 1$).

DISCUSSION

Three findings emerged from the evaluation of our on-line introductory psychology class. The course formats differed in their effects on content knowledge (as measured by multiple-choice examinations), computer use and computer anxiety (as measured by student self-reports), and satisfaction with the course (as measured by student course evaluations). First, the students in the on-line version of the course scored higher than the students in the lecture version on periodic in-class examinations; the students in the on-line version also showed larger gains during the course on a set of questions covering a broad range of topics in psychology. Second, the students in the on-line course reported higher increases in computer use and greater decreases in computer anxiety than did the students in the lecture version of the course. Third, the lecture students expressed greater overall satisfaction with the course than did the students in the on-line course.

These trends are summarized in Table 7 for eight measures, two assessing content knowledge (pretest–posttest gain in content knowledge and scores on in-class examinations), three assessing use of and attitudes toward computers (pretest–posttest gain in computer use and computer enthusiasm, and pretest–posttest decreases in computer anxiety), and three tapping overall satisfaction (interest in psychology, willingness to recommend the course to a friend, and likelihood of signing up for the same section again). For each measure, the effect of instructional method is indicated as significant or not, and, if significant, the direction of the effect is indicated (as favoring the on-line or the lecture format). The table thus shows the trends mentioned above, in that the first five measures show significant effects favoring the on-line format but the three satisfaction measures show significant effects favoring the lecture method.

Table 7 applies the same scheme to summarizing the modulation of the effects of instructional method by three other variables—instructor, semester, and class standing. None of these variables modulated differences in the content knowledge measures that favored the on-line method. The effects of instructional method on computer use, computer enthusiasm, and computer anxiety were modulated somewhat by experience; freshmen showed the most decrease in computer anxiety, and increases in computer enthusiasm were greater during the fall than during the spring semester. A different picture emerged for the satisfaction measures. The effect of method of instruction on all three measures of satisfaction was modulated by the instructor variable, and two of the measures were also influenced by the time of year (fall vs. spring semesters).

Measures of learning and satisfaction also showed different patterns when activities in the class meetings for the on-line group were used as the independent variable. Nonfreshmen students who participated in the spring review groups scored higher on examinations and on the final examination than did similar students in the demonstration groups. This difference was smaller and nonsignificant in the other conditions. Thus, there was some advantage on examinations for the review group procedure. Yet, when the students were asked whether they would take another on-line course, the students in the demonstration groups were more likely to agree than the students in the study/review groups. As with the on-line versus lecture comparisons, the patterns for learning and

Table 7
Summary of Effects of Method of Instruction on Learning, Computer Use, and Satisfaction

Measure	Method Effect (Superior Method)	Interactions With Method (Method × *Variable* interaction)
Gain in content knowledge	Yes (on-line)	No
Score on in-class examinations	Yes (on-line)	No
Increased computer use	Yes (on-line)	No
Increased computer enthusiasm	Yes (on-line)	Yes (semester)
Decreased computer anxiety	Yes (on-line)	Yes (class standing, $p < .06$)
Interest in psychology	Yes (lecture)	Yes (instructor)
Recommend course to a friend	Yes (lecture)	Yes (instructor and semester, $p < .06$)
Sign up for same section again	Yes (lecture)	Yes (instructor and semester)

Note—The cell entries indicate whether the effect of method (of instruction) for each measure was significant (yes or no) and, if so, which method was favored (on-line vs. lecture). Modulation of the effect of method by instructor, semester, and class standing are also included. Significance levels are $p < .05$, unless otherwise indicated.

satisfaction differed. Higher satisfaction did not mean more learning, and vice versa.

The lecture and the on-line sections had differential attrition rates. It is possible that the difference in performance on examinations and content learning was due to losing more weak students in the on-line sections. However, the fact that the students in the on-line sections were less satisfied with their course offers some argument against this, in that we would also need to argue that the presumably weaker students who dropped the on-line courses would have given it higher ratings had they remained in the class. This seems unlikely, because within the on-line sections offered in the spring semester, there was a significant correlation between exam performance and satisfaction (.42 for recommend course and .38 for sign-up again).

Assuming that the difference in amount of learning is not an artifact of selection, we can ask what aspects of the on-line course resulted in better content learning. Although the students read the same textbook material in the on-line and the lecture sections, the students in the on-line sections were required to pass mastery quizzes, and they did more interactive learning with the computerized exercises. Better learning may have resulted from the mastery requirement, from the more active learning format, or from the use of the media itself. Our design did not allow us to tease apart these factors, although we believe that there is nothing magical about the use of the media. Instead, our learning effects probably came from how we required the students to structure their study and to interact with textbook material.

One puzzling effect in our data is that the higher examination performance by the on-line group observed on the four midterm examinations was not present on the final examination. The students in the on-line sections obtained course credit in many ways besides examinations (including computer assignments, computerized quizzes, and answering questions about study materials). Many on-line students needed few points on the final examination to get the grade they sought, so they may not have studied for the cumulative portion of the examination as much as did the students in the lecture sections, who gained course credit only through examinations.

Lower satisfaction in the on-line than in the lecture sections may have occurred because the on-line sections did not have lectures presented by an enthusiastic instructor. Williams and Ceci (1997) found that *all* dimensions of course and instructor ratings were higher when the instructor used an enthusiastic style than when he was less enthusiastic, although learning was equivalent in the two conditions. In our data, the reported amount of time spent on lecture and on-line courses was roughly comparable, but the students in the on-line course tended to agree more strongly that they had to work harder in this than in other courses. A finding by Greenwald and Gillmore (1997) that lighter workloads are related to higher satisfaction may help to explain the students' lower satisfaction with the on-line course.

We replicated earlier studies comparing computer-based courses with traditional courses. As in Forsyth and Archer's (1997) comparison, students rated the computerized components of the course favorably. Our students found the on-line course to be more convenient than a lecture course, they liked having to come to class only once a week, but they were mixed on whether more lectures were needed. These results are consistent with those of Pear and Novak (1996). However, fewer of our students than of theirs would enroll in another on-line course, especially students in the study and review sections (where nonfreshmen, at least, learned more, as measured by examinations). Like Schutte (1996), we found that students in on-line sections learned more, but he reported a difference of 20%; our difference was much smaller at less than 5% on both examinations and our content questions. Schutte was teaching statistics and reported that students may have learned more in the on-line course because they formed study groups to deal with the difficult material. We organized study groups in some sections in the fall, but these were not successful, in that they failed to promote better learning and students did not like them.

We believe that we have learned some important things about on-line courses from the research reported here.

We are continuing to offer on-line introductory psychology courses, but we are focusing on individual differences that will predict students for whom the on-line course format is most appropriate. We hope to uncover cognitive and personality variables that will better predict both learning and satisfaction.

REFERENCES

BAHRICK, H. P., BAHRICK, L. E., BAHRICK, A. S., & BAHRICK, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, **4**, 316-322.

BARR, R. B., & TAGG, J. (1995, November/December). From teaching to learning: A new paradigm for undergraduate education. *Change*, **27**, 12-25.

BORK, A. (1997). The future of computers and learning. *Technological Horizons in Education*, **24**, 69-77.

COHEN, B. A., & WAUGH, G. W. (1989). Assessing computer anxiety. *Psychological Reports*, **65**, 735-738.

DEDE, C. (1996). The evolution of distance education: Emerging technologies and distributed learning. *American Journal of Distance Education*, **10**, 4-36.

DILLON, A., & GABBARD, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learning comprehension, control, and style. *Review of Educational Research*, **68**, 322-349.

EDUCATIONAL TESTING SERVICE (1994). *GRE: Practicing to take the Psychology Test* (3rd ed.). Princeton, NJ: Educational Testing Service.

FORSYTH, D. R., & ARCHER, C. R. (1997). Technologically assisted instruction and student mastery, motivation, and matriculation. *Teaching of Psychology*, **24**, 207-212.

GREENWALD, A. G., & GILLMORE, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, **89**, 743-751.

HILTZ, S. R. (1993). Correlates of learning in a virtual classroom. *International Journal of Man–Machine Studies*, **39**, 71-98.

KALAT, J. (1996). *Introduction to psychology*. Pacific Grove, CA: Brooks/Cole.

MAGNUSON-MARTINSON, S. (1995). Classroom computerization: Ambivalent attitudes and ambiguous outcomes. *Teaching Sociology*, **23**, 1-7.

MAKI, W. S., & MAKI, R. H. (1997). Learning without lectures: A case study. *IEEE Computer*, **30**, 107-108.

MAKI, W. S., & MAKI, R. H. (2000). Evaluation of a Web-based introductory psychology course: II. Contingency management to increase the use of on-line study aids. *Behavior Research Methods, Instruments, & Computers*, **32**, 240-245.

MERISOTIS, J. P., & PHIPPS, R. A. (1999, May/June). What's the difference: Outcomes of distance vs. traditional classroom-based learning. *Change*, **31**, 13-17.

PANERO, J. C., LANE, D. M., & NAPIER, H. A. (1997). The computer scale: Four dimensions of how people use computers. *Journal of Educational Computer Research*, **16**, 297-315.

PEAR, J. J., & NOVAK, M. (1996). Computer-aided personalized system of instruction: A program evaluation. *Teaching of Psychology*, **23**, 119-123.

SCHUTTE, J. G. (1996). *Virtual teaching in higher education: The new intellectual superhighway or just another traffic jam*? Available: http://www.csun.edu/sociology/virtexp.htm

SCHWARTZ, A. (1998). Tutorial: Perl, a psychologically efficient reformatting language. *Behavior Research Methods, Instruments, & Computers*, **30**, 605-609.

SPOONER, F., JORDAN, L., ALGOZZINE, B., & SPOONER, M. (1999). Student ratings of instruction in distance learning and on-campus classes. *Journal of Educational Research*, **92**, 132-140.

WALL, L., CHISTIANSEN, T., & SCHWARTZ, R. L. (1998). *Programming Perl* (2nd ed.). Sebastopol, CA: O'Reilly & Associates.

WILLIAMS, W. M., & CECI, S. J. (1997, September/October). "How'm I doing?" Problems with student ratings of instructors and courses. *Change*, **29**, 13-23.

WILSON, D. L. (1996, February 2). Brown U. offers on-line tutorials instead of lectures. *Chronicle of Higher Education*, **42**, A19.

WORTHINGTON, E. L., JR., WELSH, J. A., ARCHER, C. R., MINDES, E. J., & FORSYTH, D. R. (1996). Computer-assisted instruction as a supplement to lectures in an introductory psychology class. *Teaching of Psychology*, **23**, 175-181.

NOTES

1. Although we published a preliminary report of our experiences and included some tips to potential WWW-based instructors (Maki & Maki, 1997), the present report focuses on a better controlled evaluation of a more recent version of our course.

2. A significance level of $p < .05$ was used for all analyses.

3. We did analyze the spring data, including freshmen versus other. In those satisfaction data, method of instruction did not interact with college classification.

(Manuscript received November 4, 1999;
accepted for publication January 26, 2000.)