INSTRUMENTATION & TECHNIQUES

A computer-based Personalized System of Instruction course in applied behavior analysis

JOHN CROSBIE and GLENN KELLY
Deakin University, Geelong, Victoria, Australia

Microcomputer laboratories are becoming more common in psychology departments because they provide an alternative instructional medium and a resource to help combat decreasing funding and increasing student:teacher ratios. By using microcomputers to perform the testing, we were able to conduct a Personalized System of Instruction course for 51 students, without the five proctors that would normally be required. Students had low rates of procrastination and positive attitudes toward the course, and none of the students dropped out. This paper describes how the course was organized and the hardware and software employed. Difficulties arising in the course and how they might be overcome are also discussed.

A few years ago, we decided to develop a new undergraduate course in applied behavior analysis. We wanted to use the best educational techniques given our resources, and we therefore developed the course with two main components: (1) it was based on Keller's Personalized System of Instruction (PSI; Keller, 1968); and (2) microcomputers performed much of the work traditionally done by student proctors, which were not available at our university. This paper describes how computers can be used to run such a course and, hence, complements recent examples of computer assistance in teaching methodology (Cohen, 1991; Perone, 1991).

Personalized System of Instruction

Although there are numerous variations, a typical PSI course has seven components: (1) clear study objectives are provided, (2) material to be learned is divided into small units, (3) material is presented in a written format, (4) students proceed at their own pace, (5) each unit must be mastered before a student can proceed, (6) immediate feedback is provided on each unit test, and (7) proctors control testing and course administration (Williams, 1976). It has been consistently shown that PSI is superior to traditional lecture-based instruction (C.-L. C. Kulik, J. A. Kulik, & Bangert-Drowns, 1990; J. A. Kulik, C.-L. Kulik, & Carmichael, 1974; J. A. Kulik, C.-L. C. Kulik, & Cohen, 1979; J. A. Kulik, C.-L. C. Kulik, & Smith, 1976). PSI increases college-level examination scores by

an average of .48 standard deviations, which represents a shift in student performance from the 50th to the 70th percentile (C.-L. C. Kulik et al., 1990). This increased performance is also maintained, and students report more positive attitudes toward the content and instruction.

Proctors

A PSI course is controlled by an instructor who selects study materials, constructs tests and exams, and performs the final evaluation of student progress. The instructor may also present some lectures and demonstrations, but typically he/she has minimal contact with students; they interact primarily with proctors. Proctors are students who have previously completed the course and provide teaching assistance to subsequent classes for course credit and/or payment. They provide students with all the study materials (except texts), and they administer, correct, and discuss unit tests. The use of proctors is a consistent practice in PSI courses (Keller, 1974), and one proctor is typically provided for each 10 students (Keller, 1968, 1982). Green (1974) maintained that once student numbers exceeded 15, it is a mistake to attempt a PSI course without proctors.

Although it has traditionally been assumed that "the use of proctors is essential to the success of personalized instruction" (Farmer, Lachter, Blaustein, & Cole, 1972, p. 401), empirical investigations do not support this view. For example, Croft, Johnson, Berger, and Zlotlow (1976) compared achievement levels and rates of progress across four groups: no monitoring, self-monitoring (i.e., graphing of performance), weekly proctor monitoring, and bi-weekly proctor monitoring. No group differences were found in achievement or in the number of sessions required to complete the course. Furthermore, several component analyses of PSI have shown that frequent quizzing,

We are grateful to Keith Miller for providing tests and advice on how to conduct the course, and to Patrick Flanagan and Simon Parker for providing helpful comments on an earlier version of this manuscript. Correspondence concerning this article or the programs described should be sent to John Crosbie, who is now at the Department of Psychology, West Virginia University, P.O. Box 6040, Morgantown, WV 26506-6040 (e-mail: Bitnet: crosbie@wvnm; Internet: crosbie@wvnm.wvnet.edu).

written objectives, and mastery criteria are most important, and that proctors are considerably less important (Caldwell et al., 1978; J. A. Kulik et al., 1976; Williams, 1976). It has also been found that proctoring does not affect student exam performance, and that less student-proctor contact may be better than more (J. A. Kulik, Jaksa, & C.-L. C. Kulik, 1978). Hence, the didactic function of proctors is not essential.

Microcomputers

"The use of computers in education is becoming nearly as common as the chalkboard" (Chaparro & Halcomb, 1990, p. 141). This opinion has been supported by a recent survey of computer use in American college psychology instruction (Anderson & Hornby, 1990) that shows that over 80% of respondents worked in a department that used computers in psychology instruction. Of the respondents, 66% indicated that their psychology department had its own microcomputer laboratory (with a modal size of 12 machines), and 55% of departments were aiming to increase the number of computers.

How effectively such resources are used is unclear. Halliday (1991) suggested that microcomputers should be a highly useful tool in psychology classes, but are greatly underutilized. Practical courses such as those recently described by Perone (1991) and Cohen (1991) are examples of a promising alternative. Such efforts are not rare, but neither are they sufficiently common. Halliday's argument extends to general classroom teaching. Computers can be effective teaching tools (Halcomb et al., 1989), but are currently not fully exploited.

The course described in this paper utilized the computer as a reliable and cost-efficient way to test, score, and keep records for a PSI course (see also Halcomb et al., 1989). We find that much of the workload that has traditionally been costly, in terms of money or personnel, can be transferred to computers. The result is an efficient course in which staff numbers, cost, paperwork, and administrative time are kept to a minimum.

METHOD

Course Content and Structure

Fifty-one students with minimal background in behavioral psychology began an 8-week course on the application of behavior analysis to everyday human behavior. One faculty member (J.C.), one graduate student (G.K.), and a laboratory containing 20 personal computers were the available teaching resources. Miller (1980) was the text for the course. This text covers four major themes (behavioral research methods, reinforcement control, stimulus control, and aversive control), and, in typical PSI fashion, is divided into 25 small units organized by topic. Each unit consists of several pages of reading followed by three alternate sets of questions that require constructed responses. Because of time restrictions, our course covered only 18 of the units: 1-3, 5, 8-14, 16-18, 20, and 22-24.

Assessment

A student's final grade was the sum of three measures: 18 small unit tests (18%), four larger review tests (72%), and one report (10%). All unit test questions were taken from the text, and review questions (which were previously unseen by students) were taken

from materials used by Miller to validate his text. Tests were scanned and converted into ASCII files with OCR software. This technique is efficient and greatly reduces typographical errors.

Unit tests. The students were required to demonstrate mastery of the material in each unit. Mastery was defined as at least 9 of the 10 answers correct on a constructed-response (fill-in-the-blank) test, which the students could take as often as necessary without penalty. For each test, there were 30 possible questions, and the program presented a random selection (without replacement within each session) of 10 of these questions. Hence, the students could take a test on more than one occasion without receiving the same questions, and multiple versions of tests were not required.

With computerized testing such as we employed, it is important to ensure that test performance is not inflated because students remember answers from one attempt at a test to the next, or are told the answers by other students. There are several reasons to believe that in our course, test performance was not subject to such inflation. Most students mastered a test on the first attempt, and nobody took more than three attempts at a unit or review test. Hence, it is most unlikely that anybody was exposed to all unit test questions. Furthermore, it is almost impossible to remember the answers to Miller's questions, because the same scenarios are used in various contexts with different answers and subtle discriminations are required. For example, a scenario concerning roommates Bill and Ben might be employed with minor variations to demonstrate differential reinforcement, stimulus control, punishment, and extinction. The only way to master a test is to understand the concepts; seeing a question on a previous attempt is unlikely to help. Similarly, it would be almost impossible for one student to tell another the answers. At best a student could report answers to the items that he/she received, but the reportee would receive a different random sample of items and have great difficulty assigning an answer to a question if he/she did not understand the concept being tested.

To deter ill-prepared students, the score received for a test was the one obtained on the first attempt; if the score was a 10, the student received 1% toward his/her overall mark; if, however, a score of 5 was obtained (for example), 0.5% would be the final mark for that test even though mastery was subsequently attained. The prerequisite for taking a test was mastery of the previous unit.

Review tests. When the students had mastered all lessons in one of the four major sections, they took a 15-item review test. These items were of the same format as those in the unit tests and were randomly selected by the program from a pool of 100 items representing all units in the section. As with the unit tests, the first score contributed to the final mark, and mastery (i.e., ≥ 13 correct answers) was required before the next unit test could be taken. Each review test was worth 15%.

Because they are commonly based on self-pacing, PSI courses often have high rates of procrastination and withdrawal (J. A. Kulik et al., 1974). We attempted to overcome these problems in our course via external pacing: deadlines were assigned to each of the four review tests, an additional 3% was awarded to students who mastered a review test by the deadline, and 1% was deducted for every day between the deadline and the day the test was mastered (Glick & Semb, 1978). The microlaboratory and course instructors were available to the students for 1 h Monday-Friday for 8 weeks.

Assignment. We wanted to obtain an additional measure of our students' knowledge of behavioral principles to validate their performance on tests. In ideal circumstances, the students would have presented a written proposal for an applied behavioral intervention, conducted the intervention, and submitted a report of their results. Because of the students' limited behavioral background, however, a compromise was reached by setting only the proposal as an assignment. Hence, they wrote a proposal describing in detail how they would conduct a behavioral intervention to achieve one of the following results in another person: (1) a reduction in cigarette

smoking, (2) a reduction in aggressive or tantrum behavior, (3) an increase in environmentally friendly behavior, or (4) a reduction in weight. Using the *Journal of Applied Behavior Analysis* as a reference source, the students described and justified (1) the recording of target behavior, (2) the behavioral procedures to be employed, (3) the single-subject design to be used, and (4) how generalization and maintenance would be obtained.

Hardware

The microlaboratory was a spacious room (located in the Psychology Department at Deakin University) used for both teaching (e.g., research methods and statistics) and research (e.g., data analysis, word processing) purposes. The laboratory had 20 Samsung S330 (IBM-PC-compatible) computers, each with 640K RAM, a 20-MB hard disk, one 5.25-in. disk drive, and a CGA monitor.

Study Room

Adjacent to the microlaboratory was a room available for study and discussion during testing times. The room had several large tables and could comfortably accommodate 20 students. Along one side of this main room were four smaller rooms capable of seating 6 people for small-group discussion. The main purpose of these areas was to give the students a formalized and functional area to study or discuss course materials both among themselves and with instructors.

Software

Each student had a key diskette that contained an executable version of the PSI program (written in Turbo Pascal 5.0) and a file that showed the next test that he/she would receive. Each computer in the laboratory had the data for each unit and review test on the hard disk. These data were coded (128 was added to the ASCII number of each character) so they could be read only with the PSI software. For security reasons, key disks were available only during class sessions and were copy protected. To take a test, a student obtained his/her key disk, inserted it in any computer in the laboratory, and typed "PSY318" (the course code). The program performed the following functions: (1) it verified that an original key disk was in Drive A; (2) it opened the coded input file and output file; (3) it randomly selected 10 items for a unit test or 15 questions for a review test, uncoded these items, then stored them in RAM; (4) it produced a black screen with a white border; and (5) it wrote "Elapsed Time," "Correct," and "Incorrect" in a light green box in the top right corner of the screen. Elapsed time was displayed continuously in this box. For each item selected, the question was displayed in the center of the screen and the student typed an answer, which was also recorded in RAM. When all items had been completed, the students scored their answers. For each item, the question, the student's answer, and the correct answer were all displayed on the screen simultaneously. The students typed "C" if an answer was correct and "I" if it was incorrect. Whenever a student scored an answer as correct, the number of correct items shown in the score box at the top of the screen was incremented by one; the number of incorrect items was similarly incremented if an answer was scored as incorrect. The student's answer, the correct answer, and the student's scoring of the answer (i.e., "C" or "I") were all written to an output file so that instructors could verify the scoring. Having students, rather than computers, score answers is useful for two reasons. Computers have great difficulty accommodating incorrect spelling and synonyms for listed answers, and scoring has an educative function because it requires students to attend carefully to the question and answer. Student scoring simplifies the technique, may improve performance, and provides instructors with considerable information about errors that are made regularly, and therefore about bad test items.

Testing

Our course used computers to perform most of the testing and paperwork. The primary function of the instructors was to provide feedback to the students about test items and general information about course content. Test administration was controlled by the students. During a testing session, the students would take their own (personally labeled) computer disk from the disk holder at the front of the microlaboratory, go to a computer, and begin testing. After a student had completed and scored a test, an instructor checked the scoring, wrote the score on a master sheet, discussed difficult questions if necessary, and, if the test had been mastered, updated the file that contained the next test to be presented. A checking program on each key disk facilitated these tasks, so checking and updating rarely took longer than 30 sec. The only paperwork required was a master sheet of students' scores and the dates when review tests were mastered.

RESULTS

Drop-Out Rates

Failing to complete the course is the most consistently negative effect of mastery programs (C.-L. C. Kulik et al., 1990), but this effect is not ubiquitous and depends on the characteristics of a particular study (J. A. Kulik et al., 1979). Our course demonstrated a high retention rate: of the 51 students beginning the course, 49 completed all unit and review tests, and 2 students completed all but the final review test.

Deadlines/Procrastination

For the four review tests, 80.4%, 64.7%, 72.6%, and 96.1% of the class, respectively, demonstrated mastery on the first attempt and, therefore, received a 3% bonus. Only 2 students (3.9%) did not complete the coursework by the final deadline. Although it is difficult to evaluate the effectiveness of the deadline contingencies without a control group, in our previous lecture-based courses on the same topic when no deadlines were employed, students never proceeded so quickly and uniformly and never obtained such a high level of mastery.

Total Scores

Unit test, review test, and assignment grades were summed to produce the total score. A total score of $\geq 90\%$ was assigned a letter grade of A, 80–89% a B, 70–79% a C, 60–69% a D, and $< 60\%$ an F; 42% of our students were awarded an A, 32% a B, 15% a C, 2% a D, and 9% an F. Such a negatively skewed distribution is typical of PSI courses (e.g., Croft et al., 1976).

Validation

Figure 1 shows a scatterplot of test and assignment grades for all students. The two measures are moderately correlated [$r(43) = .52, p < .01$], which suggests a reasonable level of concurrent validity (Murphy & Davidshofer, 1988, p. 117). This correlation probably underestimates the true level of validity, however, because it is based on a restricted range of scores (i.e., most stu-

Figure 1. A scatterplot of assignment and test grades ($N = 45$).

dents had high test scores). A more accurate validity coefficient can be obtained by correcting the correlation for restricted range (Murphy & Davidshofer, 1988, p. 79). When this was done, the validity estimate increased [$r(43) = .63, p < .01$].

Evaluation

Most students like the traditional PSI course structure (J. A. Kulik et al., 1974). Our course was somewhat different, however, because computerized management replaced much of the traditional human contact. At course completion all students provided positive course appraisals, and the consensus was that they liked the course content and the efficiency with which the course was run.

DISCUSSION

The course was a success. In 8 weeks, 51 students learned the majority of an applied behavior analysis textbook. Performance on unit and review tests was typically high and on schedule. Paperwork was restricted to one record sheet, instructors' duties were kept to a minimum, and contact time was restricted to 5 h per week. Furthermore, the students did not lament the absence of personal proctors. This supports previous research indicating that full proctor support is not necessary in a PSI course. What is required, however, for a course like ours is (1) an adequate number of computers, (2) a controlling program, (3) someone with sufficient computing expertise to overcome any difficulties, and (4) appropriately organized test materials on disk.

Ways to Improve the Course

After a student had corrected a test, an instructor ran the checking program that displayed the student's answers and the correct answers on the screen for validation. If a student wished to debate the correctness of an answer, it was necessary to refer to paper versions of tests to find the relevant question; for security reasons, only coded ver-

sions of questions were kept on the computers, and these could not be read without a special program. Future versions of the checking program will display the question, the student's answer, the correct answer, and the scoring immediately after the self-correction phase.

The students' disks contained an output file for each test that showed how many questions had been correctly completed in the test. As a backup, a master sheet also contained each student's score on each test. A useful modification to subjects' disks, which would obviate the need for a master sheet, would be a program that (1) recorded the tests that had been mastered, (2) maintained a running total score after each test, and (3) recorded how many times a subject attempted each test before mastery was achieved.

Perhaps the most important aspect of the system described here is the dedication of instructors' time to the discussion room. Because enrollments far exceeded the anticipated class size, it was common for large demands to be placed on instructors in the microlaboratory, leaving the discussion room vacant during these times. This did not impede the students' performance or progress, and only rarely did they seek help outside class time, but many students mentioned that they would have preferred more discussion time. Inherent in the self-paced course is a large variation in usage times of the microlaboratory, so matching staff numbers to student demand is impossible. A possible solution to this may be to schedule testing times to provide greater streamlining (e.g., on Monday and Wednesday only test students whose surnames begin with A-L). One instructor dedicated to the discussion room would be ideal, but would be cost-inefficient on low-demand days.

Our course was successful, and our analysis demonstrates that computers can be used as effective substitutes for proctors when personnel numbers are limited.

REFERENCES

- ANDERSON, M., & HORNBY, P. (1990). Computer use in psychology instruction. *Behavior Research Methods, Instruments, & Computers*, 22, 194-199.
- CALDWELL, E. C., BISSENETTE, K., KLISHIS, M. J., RIPLEY, M., FARUDI, P. P., HOCHSTETTER, G. T., & RADIKE, J. E. (1978). Mastery: The essential essential in PSI. *Teaching of Psychology*, 5, 59-65.
- CHAPARRO, B. S., & HALCOMB, C. G. (1990). The effects of computerized tutorial usage on course performance in general psychology. *Journal of Computer-Based Instruction*, 17, 141-146.
- COHEN, S. H. (1991). Computer-based methodology laboratories: 2. First-year graduate courses in data analysis. *Behavior Research Methods, Instruments, & Computers*, 23, 127-129.
- CROFT, R. G. F., JOHNSON, W. G., BERGER, J., & ZLOTLOW, S. F. (1976). The influence of monitoring on PSI performance. *Journal of Personalized Instruction*, 1, 28-31.
- FARMER, J., LACHTER, G. D., BLAUSTEIN, J. J., & COLE, B. K. (1972). The role of proctoring in personalized instruction. *Journal of Applied Behavior Analysis*, 5, 401-404.
- GLICK, D. M., & SEMB, G. (1978). Effects of pacing contingencies in personalized instruction: A review of the evidence. *Journal of Personalized Instruction*, 3, 36-42.
- GREEN, B. F., JR. (1974). Fifteen reasons not to use the Keller plan. In J. G. Sherman (Ed.), *PSI: Personalized System of Instruction: 41 germinal papers* (pp. 117-119). Menlo Park, CA: W. A. Benjamin.

- HALCOMB, C. G., CHATFIELD, D. C., STEWART, B. E., STOKES, M. T., CRUSE, B. H., & WEIMER, J. (1989). A computer-based instructional management system for general psychology. *Teaching of Psychology*, **16**, 148-151.
- HALLIDAY, S. (1991). Practical computing? *Psychology Software News*, **2**, 36-37.
- KELLER, F. S. (1968). "Good-bye, teacher. . . ." *Journal of Applied Behavior Analysis*, **1**, 79-89.
- KELLER, F. S. (1974). Ten years of personalized instruction. *Teaching of Psychology*, **1**, 4-7.
- KELLER, F. S. (1982). *Pedagogue's progress*. Lawrence, KS: TRI Publications.
- KULIK, C.-L. C., KULIK, J. A., & BANGERT-DROWNS, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, **60**, 265-299.
- KULIK, J. A., JAKSA, P., & KULIK, C.-L. C. (1978). Research on component features of Keller's Personalized System of Instruction. *Journal of Personalized Instruction*, **3**, 2-14.
- KULIK, J. A., KULIK, C.-L., & CARMICHAEL, K. (1974). The Keller plan in science teaching. *Science*, **183**, 379-383.
- KULIK, J. A., KULIK, C.-L. C., & COHEN, P. A. (1979). A meta-analysis of outcome studies of Keller's Personalized System of Instruction. *American Psychologist*, **34**, 307-318.
- KULIK, J. A., KULIK, C.-L. C., & SMITH, B. B. (1976). Research on the Personalized System of Instruction. *Journal of Programmed Learning & Educational Technology*, **13**, 23-30.
- MILLER, L. K. (1980). *Principles of everyday behavior analysis* (2nd ed.). Belmont, CA: Wadsworth.
- MURPHY, K. R., & DAVIDSHOFER, C. O. (1988). *Psychological testing: Principles and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- PERONE, M. (1991). Computer-based methodology laboratories: 1. An undergraduate course in experimental psychology. *Behavior Research Methods, Instruments, & Computers*, **23**, 121-126.
- WILLIAMS, R. L. (1976). Personalized System of Instruction: Future research areas. *Journal of Personalized Instruction*, **1**, 106-112.

(Manuscript received July 27, 1992;
revision accepted for publication April 13, 1993.)