Can computers help with one of the most persistent challenges of large classes: the accurate and efficient measurement of student learning?

Computerized Testing in Large Courses: A Case Study

John F. Kremer

This case study is about a highly enrolled introductory psychology course at a large (twenty-eight thousand students) research-oriented urban university. This course was taught in large lecture halls. Students’ grades were determined solely by their performance on multiple-choice tests. The “DWF rate” (those who did not pass the course) was consistently around 40 percent. More than ten years ago, we started with the hypothesis that this rate could be reduced and that these students could learn to succeed. The first step in this process was to institute computerized testing. Annually, the Psychology Testing Lab assesses over four thousand students. The lab is a four-hundred-square-foot room that originally supported sixteen computers. Students take five tests per semester, and they are permitted to take each test twice. Initially, when tests closed on the weekend, students often waited in long lines. The focus of this chapter will be on the lessons that we have learned in implementing computerized testing with limited resources.

Lessons Learned

Computerized testing in large courses provides flexibility for students and faculty as well as solving some space problems. However, test security issues can continue to be a challenge. In the final analysis, computerized testing encourages course development and coherence.

Flexibility of Computerized Testing Provides Numerous Advantages for Students and Faculty. Students may take tests almost any time of the day and any day of the week. Our urban students have myriad other life responsibilities. For example, students spend more time working (twenty-two
hours per week) than attending class (eighteen hours per week) or studying outside class (fifteen hours per week). They appreciate the ability to plan their tests around their other academic and life tasks. Flexibility also helps faculty. Faculty members can open and close the tests at their discretion and can easily extend the test for individual students. The ease of providing students with a makeup test has enabled faculty to be more responsive to the unexpected demands in students' lives.

**Space and Other Problems Can Be Solved.** Administering well over ten thousand tests per semester in a four-hundred-square-foot room with sixteen computers was a daunting task. We had long lines, complaints about noise, and actually a fight or two. Problem solving gradually reduced these problems, and the lab today has few lines and complaints. Several strategies pushed this progress:

The test closing dates for individual sections are staggered based on enrollment.
Students are required to wait twelve hours before taking tests again.
On a “Testing Reservation” home page, students can find the times when the lab is busy and reserve one of the testing computers.
If a line occurs, students are given a return time, and a computer is reserved for them.
The lab was also redesigned with smaller compact computers and screens to reduce crowding and yet increase the number of computers by 40 percent.

An additional problem was the amount of noise caused by the number of people in the small lab, noisy keyboards and computers, and student-proctor conversations. Changing to soundless keyboards and quieter computer fans, adding carpet, and providing more space between students reduced the noise. Students may also use earphones and listen to repeating sounds or melodies that mask the noise in the room and from the hall outside. To reduce student-proctor conversations, students sign into the lab on a keyboard mounted to the open door of the lab, and the computer gives directions and assigns students to a testing space.

**Test Security Issues Are a Persistent Challenge.** Initially, the lab was designed to be low maintenance: Students would simply go into the lab and take their tests with only sporadic and random supervision. However, students soon complained about the high amount of student dishonesty. Over the years, types of academic misconduct included unauthorized written notes, unauthorized notes on the computer screen, conversations with other students, hacking into the test bank, printing off the tests, and taking the test for other students. A variety of strategies were used to counteract these problems. The following steps were instituted to reduce the probability of similar items appearing on adjacent computer screens:
• Each test has five versions of nonoverlapping items
• The items for each test are randomly presented to students, and no student gets the same test twice
• Test items are randomly presented one at a time
• Work-study students are used as testing proctors
• Work-study students are trained, supervised, and monitored
• Proctors are able to easily observe student screens after the room was rearranged, the proctor’s station was elevated, and proctors were given a computer program to directly observe each student’s computer screen
• Video cameras were strategically placed in the lab
• Printing was disabled from any of the computer terminals
• A computer programmer effectively secured the test items from hackers
• To prevent one student from taking the test for another student, proctors check the students’ identification cards
• Students swipe their identification card to gain access to the test; the test then is available only to the student who is identified on the card
• To prevent many different types of academic misconduct, students may hit a button on the computer screen to report cheating to a proctor in the lab
• Before taking the test, students must read a code-of-conduct statement and agree to abide by the principles in the statement; this promise includes the responsibility to report incidents of academic misconduct when they occur

By having a central testing location, most of the unethical behaviors that occur not only in the lab but also in a typical classroom are eliminated or reduced. Computerized testing presents unique unethical possibilities, but these are easily identified and reduced by careful problem solving.

**Computerized Testing Encourages Course Development and Coherence.** Centralizing the testing function of a large course frees up resources to develop not only the administration of the test (items noted above) but also the content of the test and other course components. The following strategies helped integrate the many components of the course into a more cohesive package.

**Different Types of Test Items and Item Responses Were Added to Better Measure Course Objectives.** After initial programming, one of the principal advantages of computer assessment is the ability to easily combine several different methods of assessment. One of the objectives for this course was to help students apply psychological concepts and theories to real-life examples. So test items include not only written applications but also audio-visual items from existing movies or acted scripts that were taped for the test. On the response side, the tests combine traditional multiple-choice answers and short-essay questions. To reduce faculty time in scoring the twenty-five thousand essays, student raters grade the essays using scoring
criteria and sample essays to enhance reliability. Raters have access to the essays from any Web site and typically grade the essays within twenty-four hours. However, with this volume of work, raters are not able to provide adequate feedback. A computerized essay content grader will soon be used to not only grade essays but also give feedback to students.

Content Validity of the Test Was Improved. Because the same five tests are used across forty sections of this course, a significant amount of time and attention is given to the development of items. For example, this course has five critical thinking goals based on Bloom's (1956) taxonomy. Each chapter has a grid with about one hundred fifty cells (thirty concepts, names, or theories by five critical thinking skills). The cells of the grid are used to develop test questions and items for the computer exercises and written homework. The grid is also available to students and faculty to help them prepare for tests and plan class activities.

Focus of Computerized Activities Was Expanded to Include Not Only Evaluative (Summative) Assessment but Also Developmental (Formative) Assessment. When the tests were working both administratively and conceptually, resources were directed toward developing homework assignments and computer exercises to help students learn the material and give them developmental feedback when they need it and based on their performance. Separate exercises focused on each of the critical thinking goals of the course, and other exercises helped students put it all together and review for the tests.

Summary and Conclusion

I hope that this chapter makes a case for using computerized testing in large courses by pointing out the following advantages: Multiple sources of assessment can be easily integrated, audiovisual items and multiple-response formats can be used to measure higher-order thinking skills (application, analysis, and synthesis), students' educational needs can determine the timing and type of feedback, instructors can easily assess students' progress toward their learning goals, variability in assessment across instructors and raters is reduced, and assessment goals can be integrated into all aspects of the course. On the other hand, computerized testing does come with its share of hassles (too little space, crowded conditions, new test security issues). An increase in the use of technology creates questions and needs that professors are not accustomed to answering. By combining sections and courses and reorganizing resources, departments and schools can engage in strategic planning to meet these needs. The use of computerized testing and other efficient uses of technology can provide better services to students while decreasing the load on instructors and making a better use of physical resources. However, departments or schools must make a commitment to stay abreast of the newest developments in technology to improve and to even maintain a computerized testing program.
Reference

JOHN F. KREMER is professor of psychology at Indiana University–Purdue University Indianapolis.