EDITORIAL

Evaluation is a topic of considerable interest and importance in distance education. Most discussions of evaluation have focused on assessments at the program or course level, or have looked at student achievement outcomes. In this issue of DEOSNEWS, Hoi Suen and Jay Parkes take a close look at the process and measures of performance assessment. They relate the characteristics of different performance measures to the unique characteristics of the distance education context in order to make recommendations about appropriate ways to evaluate students in this context.

CHALLENGES AND OPPORTUNITIES IN DISTANCE EDUCATION EVALUATION

Hoi K. Suen (HKS1@PSUVM.PSU.EDU)
Jay Parkes
Department of Educational and School Psychology
The Pennsylvania State University

INTRODUCTION

The two disciplines of distance education and educational assessment have seen dramatic changes and growth recently. Although the two areas have been developing concurrently, they have been doing so rather independently. For example, technology has been and will continue to be a boon to both distance education and educational assessment. Computer networking is opening the way to virtual classrooms and unprecedented communication (Myrdal, 1994); while in assessment, computers have been used for such diverse purposes as record-keeping and item-banking (Wainer, Dorans, Green, Flaugher, Mislevy, Steinberg, and Thissen, 1990).

These are just two examples of the parallel development occurring. What has gone unmentioned is how the two sets of development could, and eventually will, interact. Some of the characteristics that make distance education unique will affect how assessment innovations are employed in distance education. These interactions need to be mentioned because they will affect future practice and research in assessment in distance education.

SOME UNIQUE CHARACTERISTICS OF DISTANCE EDUCATION

The most unique feature of distance education, not surprisingly, is distance. This has a myriad of implications for the learning and assessment processes. Among these implications are the isolation of
the learner from resources, support, and peers; the lack of face-to-face interaction with instructors; and delayed feedback. These factors have, in part, necessitated a larger quantity and diversity of media and technology, which becomes a second distinguishing feature of distance education.

Cropley and Kahl (1983) detail some of the aspects of this isolation. In a face-to-face education setting, the learner is surrounded by other learners and an atmosphere focused on learning: from desks, books, chalkboards, and other students to libraries, student activity centers, and tutoring facilities. Often, the distance learner is in a setting meant primarily for other purposes such as the home, or the workplace. Kahl and Cropley (1986) report distance learners more often face household duties or have full or part-time jobs than students in a face-to-face setting. The lack of contact with other students can have a significant effect on the learner's motivation. The distance learner lacks the social aspects of learning, such as competitiveness, fear of public failure, or peer pressure to conform (Cropley and Kahl, 1983). There are rarely "class clowns," "curve busters," or "teachers pets" in distance education. This interaction with others in a learning focused environment may not at first appear to be extremely important. However, other learners and the environment can greatly affect the learner's motivation for learning. This shifts a considerable amount of responsibility to the student to deal with and compensate for this lack of contact (Cropley and Kahl, 1983). It makes necessary that which in a fact-to-face setting is merely optional: the active learner. Although Kahl and Cropley (1986) found in one study that distance learners tend to have clearer, more differentiated reasons for enrolling in a particular course, Beaudoin (1990) notes distance learners are at varying degrees of readiness to take on this responsibility. This sets up somewhat of a paradox: each student thus needs some degree of personalization or individualization of instruction to suit his or her particular situation while the distance and lack of instantaneous communication cause delivery systems to be formal and rigid (Granger, 1990; Beaudoin, 1990). Distance educators have begun to address this paradox through technology.

Originally, technology was employed in distance education to provide some equity in information access (Wagner, 1990). In its more traditional forms -- such as workbooks, audio tapes, videotapes, radio and television broadcasting -- it was hardly interactive. These media are unable to address the issue of personalized learning, and in fact, are the formal and rigid forms mentioned earlier. Other forms of media, such as audioconferencing, teleconferencing, and facsimile transmission are more interactive and thus better address the needs of the distance learner. Lately, such possibilities as e-mail, computer networks, and Internet access have made distance education a "virtual" mirror image of face-to-face education with, perhaps, even some improvement.

The computer has made the cyberspace classroom possible. In such a classroom, discussions can be held, assignments given and received, questions asked and answered, collaborative projects assigned and completed, and so on. The advantage over a traditional classroom is that all of the participants do not have to be physically present at the same place or even the same time. Students, tutors, and faculty can "log-on" and engage in all of the
above activities as their schedules will permit. One of the elements missing still is the non-verbal communication (Barnes, 1995): the confused look, the attentive posture, the "light bulb" coming on. Notwithstanding, such an approach would go a long way to addressing the needs of distance learners.

Part of the difficulty in computer communication at present is the availability of computers, knowledge of computers, and knowledge of networking. Myrdal (1994) noted that in Iceland, for instance, when implementing on-line education, over 90 percent of faculty and students had never used the computer networking capabilities prior to the attempt reported. Another attempt to include e-mail transactions in a course for teachers in California encountered students who had no access to computers or little knowledge about computers (Fisher and Desberg, 1995). A project in Canada found that even when students had access to computers and knowledge about them generally, students still needed considerable time to master the techniques involved in using e-mail for discussions and communications, and searching and retrieving information from remote sites (Barnes, 1995).

Despite these challenges, computer networking holds considerable promise for the distance educator. Given this new form of instructional communication, assessment in distance education, too, will have to adapt to it.

CURRENT EDUCATIONAL ASSESSMENT APPROACHES

In the area of educational assessment, a number of fundamental innovations and changes have also occurred. From the 1940s through the 1970s, formal educational assessment had been dominated by objective testing, epitomized by such assessment formats as multiple choice testing with optically scannable responses. Over the past 15 years, several innovations and changes in emphasis have occurred. These include the development of computer-assisted technology and the increased use of alternative assessment approaches for formal student assessments. Today, a large array of assessment approaches are available to the educator.

Alternative Assessments

One of the major changes over the past 15 years is the increased advocacy or actual use of alternative subjective approaches for the formal assessment of student ability or achievement. Two major categories of these alternative assessment approaches have emerged: authentic/performance assessment and portfolio assessment. Advocates for these approaches have argued that these approaches are inherently superior to objective multiple-choice testing.

The major goal of authentic performance assessment is to assess the ability to apply knowledge to solve real-life problems. Baker, O'Neil, and Linn (1993) listed the following six characteristics of a performance assessment: It 1) uses open-ended tasks; 2) focuses on higher order skills; 3) employs context sensitive strategies; 4) often uses complex problems requiring several types of performance and significant student time; 5) consists of either individual or group performance; and 6) may involve a significant degree of student
choice. These types of assessments, at least insofar as a general description is concerned, approach the learner as more active. The student must take considerable control over the assessment through planning and applying knowledge in perhaps new and different ways. Proponents of these methods claim, too, that they reach more complex cognitive skills (Wiggins, 1989).

Reckase (1995) defined a portfolio as a purposeful collection of student work that exhibits to the student and others the student's efforts, progress, or achievement in a given area. This collection must include 1) student participation in selection of portfolio content, 2) the criteria for selection, 3) the criteria for judging merit, and 4) evidence of student self-reflection. Portfolios, even more so than other forms of performance assessment, call on the learner to be highly involved in planning the entries, choosing what to include, and providing the rationale behind those decisions. Portfolios thus attempt not only to assess the end products, but to some extent, the process that went into creating them as well.

Computerized Assessment

Parallel to the developments in alternative assessments, advances in computer technology have also brought about some developments in objective testing as well as providing the potential for other innovations in alternative assessments. With the common availability of high-speed computer technology, objective testing has been made more efficient through a number of different designs. One simple design is computer-assisted testing. With this design, a conventional objective test is administered on screen rather than through paper and pencil. The advantage of this design is the efficiency in scoring and report generation.

A slightly more sophisticated design is the generation of objective testing through item banking. In this approach, the computer is used as a repository of numerous objective test items with known statistical properties. Tests can then be generated through a computerized selection of items that will meet certain content and statistical property specifications. The advantage of item banking is the ability to generate tests that are customized to the need of each instructor. Additionally, it is possible to develop algorithms such that examinees will respond to comparable but not identical tests. This maximizes test security and scheduling flexibility.

A third design is computerized adaptive testing, which is also known as tailored testing. Through this design, after each response to each computer-administered item, a program estimates the examinee's ability based on the examinee's responses to all previous items. The program then selects for the next item one with the difficulty which best matches the examinee's estimated ability. The process is then repeated iteratively until some criterion, such as a preset level of score precision, is met. This is the most efficient of all objective testing approaches and generally requires only about one-fifth the number of items otherwise needed in conventional paper-and-pencil objective tests to attain a given level of score precision.

Finally, computers have also been used to approximate authentic performance assessment. This is typically accomplished by employing multimedia technology to simulate authentic problem
contexts. The examinees are then to either take the proper actions to respond to the problem or to select the proper response from a number of available alternatives.

STUDENT ASSESSMENTS IN DISTANCE EDUCATION

With the advances in student assessments, an apparent question is whether distance education can benefit from any of these changes. Also, with the large variety of available assessment approaches, are there any specific approaches that might be particularly suitable for student assessment in distance education? Since student assessments serve many different purposes in the learning process -- such as placement decisions, formative evaluation, diagnostic evaluation, and summative evaluation -- the answer is not straightforward. These purposes apply to both traditional and distance education. In light of some of the unique characteristics of distance education, however, certain assessment approaches might be quite suitable for some purposes but not others. We will examine each of the assessment approaches to evaluate their suitability for each of the purposes in distance education.

Conventional Objective Testing

The conventional paper-and-pencil test, typified by a multiple-choice test, is one of the most widespread formats of assessment in education. The main advantages of this type of assessment is the efficiency and economy of administration and scoring. It is also relatively easy to construct such tests with highly reliable scores. However, such advantages of efficiency and reliability are primarily gained through the ability to test many students simultaneously under the same standardized, controlled setting. When applied to a distance education setting, requiring all students in a class to be tested through the same standardized procedure at the same controlled location is not feasible. If we were to adjust to the distance education environment by administering conventional objective tests individually at the students' own location, the advantages of efficiency and reliability disappear. Students would take the exam at different times under different conditions with different available resources. Text security is nonexistent. We cannot even ascertain that the responses to the test actually came from the student. These factors would probably render the scores from such assessments insufficiently reliable for important decisions.

Therefore, the appropriateness of conventional multiple-choice testing for student assessments in distance education may be quite limited. Specifically, such testing would not be appropriate for placement decisions and summative evaluations. The use of such tests for these purposes would be justified only if students can be tested at designated test centers under controlled settings. This would essentially be a system of distance learning but centralized assessment.

The conventional objective testing procedure, however, might be useful for the purposes of formative evaluation and diagnostic evaluation, provided that we recognize that the results of each
assessment are tentative, pending additional evidence. For the purpose of formative evaluation, tests can be developed along with instructional materials. This will work well with the self-management required of distance education students. Students can work with materials until ready for the test and take it when they are ready. Successful completion of the text provides indication that the student is ready for the next unit of instruction. Reliability is not a major concern in this case because if the test scores are in error, the student would experience difficulty in the next unit. In such a situation, the student can simply return to the previous unit for reviews.

Computer Assisted Testing

With computer-assisted testing, the item banking approach is used. Objective test items are stored in a central computer server. As needed, various equivalent versions of the same test can be generated through such procedures as random selection of items or selection of items based on a specified mix of difficulty level and discrimination power. This approach to assessment is not very different from the conventional paper-and-pencil objective test. Items are presented on screen instead of on a piece of paper and students respond by making selections on screen rather than darkening an optical scanner sheet.

The disadvantages of this approach are quite similar to those of the conventional objective test. However, there are a number of additional disadvantages. First, when implemented, this approach would require that all students have access to a computer linked to the central server. There also needs to be a large item pool with known difficulties and discrimination power generated in stores on the server. Therefore, a first requirement is that all students have access to the necessary technology, which may be problematic (e.g., Beaudoin, 1990). The instructor also needs to be knowledgeable of both the computer technology and test item analysis in order to generate, store, and manage the item bank. Finally, this approach can be expected to be more costly than conventional objective testing.

The difficulty of implementing conventional objective tests in distance education to maintain the advantages of efficiency and reliability also applies to computer-assisted testing. The inability to maintain high reliability, however, is reduced for computer-assisted testing. Specifically, the test security problem is alleviated in that different students will be administered equivalent but not identical tests. Additionally, it would not be possible for anyone to obtain a copy of the test because there is only a large item bank, not a test. In spite of this advantage, however, there is still no control over who exactly responds to the test and what resources are available at the time of the test. Therefore, computer-assisted testing is also not suitable for placement and summative evaluations. For formative evaluation, however, computer-assisted testing does have an additional advantage over the conventional objective test. Students can take the test many times until they can demonstrate mastery. At each administration, an equivalent but not identical test is used. This is not possible with the conventional objective test. Finally, as computer networking becomes more feasible, it would be possible for students to log on at any time to do formative or diagnostic testing, adding flexibility as well as additional support for distance
learners.

Computerized Adaptive Testing

This approach shares many of the advantages and disadvantages of computer-assisted testing. The only advantage of this approach beyond those of computer-assisted testing is that of efficiency in that a student needs to respond to substantially fewer items to attain a score with a given level of reliability. Since we cannot control the testing environment in distance education, such an advantage would be moot as reliability is not guaranteed.

This approach is generally not practical for a typical distance education setting. In order to employ computerized adaptive tests, item response theoretic techniques must be used to analyze the items in the item pool, and examinee scores are estimated through an optimization algorithm. To accomplish these tasks, large samples of subjects and items are needed initially to calibrate the items. Dependent upon the exact mathematical model employed, the number of subjects needed for the initial calibration during the development of an item pool for a single subject ranges from 200 to 1,000.

Simple Performance Assessment: Essay Exams

Conventional objective testing, computer assisted testing, and computerized adaptive testing share a common limitation. Specifically, it is typically difficult and time consuming to construct test items that are capable of assessing higher-order problem-solving skills. Within Bloom's taxonomy of cognitive levels, these tests are typically used to assess knowledge and comprehension and rarely used to assess application, analysis, synthesis, or evaluation. When the goal of an assessment is to assess higher-order cognitive skills, authentic performance assessment has been most frequently recommended.

The essay exam represents a form of assessment that is capable of assessing higher-order skills and may, thus, be considered a form of performance assessment. However, essay exams may be an authentic performance assessment or a proxy of such an assessment. When the essay exam is used to assess a student's knowledge in some content domain, it is a proxy in that it lacks authenticity.

The advantages and disadvantages of using an essay exam in a distance education setting are similar to those of using a "take home" exam in the conventional classroom. That is, the essay exam is capable of assessing the depth of knowledge on a particular topic. However, this assessment of depth is gained at the expense of assessing the breadth of knowledge. Reliability of scores are typically low for these assessments, and can be expected to be even lower than the use of objective tests for distance education. One of the reasons for an expected low reliability is the non-standardized assessment environment in distance education. This problem is common between objective testing and essay exams in distance education. However, essay exams contain an additional source of measurement error, i.e., rater bias and random rater error. Test security for essay exams in distance education is no more of a concern than it is for a take home exam in the conventional
classroom.

Therefore, as with conventional objective tests and computer assisted testing, essay exams can be quite appropriate for formative and diagnostic evaluations in distance education. However, similar to the situation with the other two assessment approaches, essay exams are not appropriate for high-stakes summative or placement evaluations because of the difficulties involved in obtaining adequate precision or reliability for such purposes.

Complex Authentic Performance Assessment

The typical authentic performance assessment involves the assignment of problem-solving projects for which students are given an extended period of time to work. There is no single correct answer and there may be many different ways of solving the same problem. Students will need to apply knowledge across domains and use various available resources. The student is evaluated on both the product and the process of the task. That is, the student is evaluated on both the outcome of the performance and on how the student arrived at that outcome. Typically, the student is required to keep a journal describing his/her approach to the problem, getting resources, and solving the problem.

As authentic performance assessments do not require any particular control of the assessment setting nor is the process to arrive at solutions standardized, they appear to be particularly suitable for distance education. The inherent flexibility of authentic performance assessment is highly compatible with the unconventional learning environment of the distance education student. In terms of reliability, however, authentic performance assessment shares the same limitations as those of the essay exam. Thus, it is quite appropriate for formative and diagnostic evaluations but quite problematic for summative and placement evaluations.

An important feature of authentic performance assessment in the conventional classroom is its flexibility in accommodating cooperative learning. An authentic performance task may be assigned to a team of students as a team project rather than to an individual. Through such team projects, we can assess product, process, attitude, and team work. Unfortunately, because of the inherent isolation of the distance learner, this feature is not practical for distance education. Therefore, the use of authentic performance assessment in distance education will most likely be limited to the assessment of an individual student’s problem-solving skills, but not effects in the area of team work. Again, the promise of computer networking may, at some point in the future, overcome this limitation.

Portfolio Assessment

A portfolio assessment appears to be quite appropriate for distance education. With this approach, the student and the instructor discuss prior to instruction the criteria for evaluation and determine how and what should constitute the portfolio. This portfolio may contain writing samples, reports of authentic performance tasks, conventional objective testing results, and/or computer-assisted objective testing results. The instructor and the student together
determine the appropriateness of each piece of evidence in the portfolio and the rating rubrics to be used. The student can then submit the portfolio at the end of the course for instructor rating.

For distance education, portfolio assessment appears to be the ideal approach for summative evaluation. Even though the results of the conventional objective test, computer-assisted test, essay exams, and/or authentic performance tasks in the portfolio are individually unreliable, the rating of the collective portfolio, however, can be expected to be much more reliable. In other words, through this approach, reliability is built in through the size of the sample of performance items and tests. This approach would provide the most reliable information for summative evaluation.

Caution must be taken, however, since the cumulative reliability may still not be very high by conventional reliability standards, making high-stakes decisions problematic. Of all the approaches, portfolio assessment holds the promise for the maximum attainable reliability. As Thorndike & Hagen (1969, p. 194) stated and reiterated later (Thorndike, Cunningham, Thorndike & Hagen, 1991, pp. 109-110), "if we must make some decision or take some course of action with respect to an individual, we will do so in terms of the best information we have, however unreliable it may be, provided only that the reliability is better than zero, in which case we have no information." Therefore, if we must make summative and placement evaluation decisions, we seek to use the most reliable information available. In distance education, outcomes of portfolio assessment appear to provide such information.

SUMMARY AND DISCUSSION

The unique characteristics of distance education pose certain challenges to the process of student assessment. The usefulness of many of the assessment approaches available for the conventional classroom is limited for distance education because of the lack of control of assessment conditions, the unique set of available resources, and the inherent isolation of the distance learner. Computerized adaptive testing and group authentic performance assessment can be considered impractical for distance education. Conventional objective testing and computer assisted testing may be used for low-stakes formative and diagnostic evaluations for low-order cognitive skills. Essay exams and individual authentic performance assessment may be used for low-stakes formative and diagnostic evaluations of high-order cognitive skills. For high-stakes summative and/or placement evaluation, portfolio assessment appears to be the only justifiable approach. This is not because portfolio assessment outcomes are highly reliable. Rather, given the available alternatives for distance education, portfolio assessment outcomes hold the promise of providing the most reliable information. If the assessment decision is extremely important and a very high level of score precision is required, portfolio assessment outcomes may not have an adequate absolute level of reliability to be defensible. For these situations, distance learners should be assessed at some centralized location under a controlled assessment setting.

REFERENCES


