COMPUTERIZED ASSESSMENT TO ENHANCE LEARNING

Lim Tock Keng

COMPUTERIZED ASSESSMENT TO ENHANCE LEARNING

Tock Keng Lim
National Institute of Education
Nanyang Technological University
469 Bukit Timah Road, Singapore 259756
E-mail: LIMTK@NIEVAX.NIE.AC.SG

Abstract

The current availability of cheap but powerful microcomputers has made test delivery by computer both feasible and attractive. It has brought about innovations in computerized testing to include adaptive testing, diagnostic testing and test banking facilities. This paper shows how a computerized adaptive test (CAT), which can administer and score a test over the computer as well as tailor the test to the examinee's ability, can be used to enhance learning. As the CAT procedure tailors the test to the examinee's estimated ability level, diagnostic features to enhance learning can be built into the procedure to allow the examinee to go through supplementary exercises on the skill he is deficient on, before returning to the test. Diagnostic feedback such as the ability to delay re-testing interval as well as realistic simulation situations can also be introduced in the CAT. The above features to enhance learning through assessment can be introduced into a local CAT in reading comprehension for Grades 3, 5, 7 and 9, currently being developed in the National Institute of Education.

Introduction

Computers have been utilised in mathematics, science and language testing in the last two decades, mainly in the analysis of test results, generation of items from item banks, automatic scoring of test items, and immediate knowledge of results. Test items have been, as pointed out by Benjamin (1988), a key element of computer-based learning virtually from its inception. The current availability of cheap but powerful microcomputers has made test delivery by computer both feasible and attractive. As shown in many studies (Alderson 1991; Anandam, 1994; Canale, 1986; Jacobson, 1993, Kumar & Helgeson, 1995), there is much work in the U.K and the U.S. on computer-based tests (CBT) and computerized adaptive test (CAT).

CBT refers to tests delivered and scored by computers, with language testing being initiated in the U.K. in the 1980s with projects such as that established in the University of Lancaster by Alderson. In addition to administering and scoring tests using the computer, CAT tailors or adapts the test to the person's ability to allow him to have an assessment of his ability with a shorter test. CAT is a computer-assisted interactive process which facilitates the rapid, accurate measurement of the ability of the testee. CAT began mainly in the U.S., through the exploration of large-scale testing by the military (Gorman, 1977; Ree, 1977) and is currently used both in the universities as well as in national standardised testing.

According to Ring (1992) and Harvey and Reid (1994), research indicates that students prefer computerized testing over conventional testing, provided that the testing environment and the test itself are both well designed. Harvey and Reid further believed that testing also plays a key role in the acquisition and learning process. The focus of this paper is on CAT, on how adaptive testing can enhance learning of the subject being assessed.

Computerized Adaptive Testing (CAT)

CAT makes use of the Item Response Theory (IRT) in measurement, a relatively new test
theory developed over the past forty years. IRT defines the relationship between an ability trait being measured and observable performance on the test. It predicts the probability of a correct response to a test item as a function of the test taker's ability and certain characteristics of the item, one of which is ability. If the person is much more proficient than the item, then this probability is large. We learn the most when the expected probability is close to one-half. An examinee's proficiency is calculated from the difficulty of the items that are presented to him.

The real strength of IRT, according to Wainer (1990), is in its dealing with items one-at-a-time. IRT posits an underlying, unobserved trait, on which the items are linearly arrayed from the simplest to the most difficult. The goal of testing is to be able to array the examinees on the same continuum as the items, from novice to expert. This goal meant that the examinee need only be presented with only enough items (rather than with all the items) for the examiner to accurately place the examinee on the latent continuum. The capacity to rank all the examinees on the same continuum gives rise to testing that can be individually tailored to each examinee. Essentially, the availability of IRT as well as computers in education has brought about the adaptive testing in CAT which can be tailored to each examinee’s ability.

Figure 1 shows a flow chart on the on-line administration of an adaptive test, a reading comprehension computerized adaptive test of a CAT project in the National Institute of Education (Lim, Ho and Wong, 1995). The student, seated at the computer terminal, is presented with a test item in the middle of the prospective ability range. If it is answered correctly, the next question is
more difficult. After each response, an estimate of the student's ability is given and updated as the testing proceeds. Essentially, the testing procedure "adapts" the test to the student on the basis of the response patterns by presenting items successively more appropriate to the individual's (estimated) ability level. Items are given until the student's proficiency is established within a predetermined level of accuracy.

Enhancement of Learning

Currently, teachers can use testing in an instructional way given the right kind of activities and programs. Willets (1992) maintained that computer-assisted teaching provided a comprehensive, fast and accurate way to test language skills. Students could also self-test using computer-assisted instructional (CAI) programmes. In science assessment, Kumar and Hegelson (1995) felt that the current technology of using computerized administration of multiple choice tests drawn from item banks could be extended to include strategies such as giving immediate feedback to students, formative evaluation with remediation possibilities, and monitoring of homework.

In adaptive testing, one consequence of targeting items to the ability level of the examinee is that examinees of different ability levels may be presented with items in different difficulty order. A low-ability examinee will generally not be able to answer the initial items correctly and will then receive easier items. The immediate feedback on the on-going performance of the examinee to such a test means that a computer-based diagnostic system can be integrated into the adaptive test as patterns of responses of examinees to the test would reveal detailed information about the strengths and weaknesses of the examinee. In the local CAT project in reading comprehension, the MCATL language of MICROCAT (the software used) will be utilized to explore the implementation of a diagnostic system using the custom interface included in MICROCAT to link FORTRAN or Pascal procedures to MICROCAT. This diagnostic system built into the Rasch analysis will allow students to be immediately aware of the errors once made they have made and their skill (ability) level attained.

At the simplest level for interaction with the learner for diagnostic purposes in the CAT, a HELP facility can be introduced to the CAT test where the learner is able to self-diagnose their errors and request for help. The HELP facility could have many levels allowing the learner to find meanings and other information of words and even pronunciation. As pointed out by Alderson (1988), a generous portion of menus and associated support (e.g., varying test with or without HELP, varying test topics, skill levels and linguistic complexity) would allow students to decide what they want to be tested and the level of difficulty. Such decisions by students and the level of HELP requested could be entered into their personal files to allow the teacher to monitor their progress.

For diagnostic and remedial purposes, further computer-based learning exercises on comprehension skills can be built into the CAT test. Students who are given feedback on their deficiency in a particular skill will be able to work on the supplementary exercise and return to the test to demonstrate their competence in the skill. In this way, examinees will be able to use the CAT reading comprehension test as a self-test as well as improve on what comprehension skill they were deficient in.

Harvey and Reid (1994) proposed the inclusion of an "expanding rehearsal" hypothesis into the diagnostic feedback given in a computer-based test. If an item is answered poorly, it should be re-presented shortly after the initial presentation until while the student is able to recall the model answer from short-term memory. If it is answered correctly the next test of that item is delayed by some time. This cycle continues so that the re-testing interval, ideally, gradually expands - the student is led to recall the answer over longer and longer intervals. The MCATL language of the local CAT test on reading comprehension can be written to take in this feature of diagnostic testing.

Paper and pencil tests usually have a fixed format as all details of a given question must be provided at the same time and items need to be free of ambiguities and errors. However, Harvey and Reid argued that in a real life situation, problems are rarely well defined. In a computer version of a tests there are opportunities to set up realistic simulation situations where some details can be given at a later stage of a problem. In a question on how a trainee teacher might cope with a classroom situation, it is possible to withhold certain details, such as student record file, until the trainee
constructs his responses. It is possible to find out how trainee teachers are able to make judgements, whether hasty or well thought out. Ring (1992) also supported the use of constructed response items based on realistic simulation situations which are close to what they are going to face in their jobs. The face validity of such items would be enhanced and provide better learning opportunities for the student.

The computer can improve assessment procedures by providing learning situations in a testing session. The adaptive testing and learning process can eventually bring about a self-optimizing learning environment where both the difficulty of the materials and the ability of learners can be empirically, progressively and automatically estimated by collating learners’ responses to learning tasks (Jones, 1994). The learners are able to interact with the computer on a one-to-one basis when they need to. In such a system, assessment happens automatically too, as a by-product of learning. In Singapore, the local CAT test of reading comprehension will be extended to include the learning features and simulations discussed and eventually to bring about the self-optimizing learning environment.

REFERENCES


Title of Submission: COMPUTERIZED ASSESSMENT TO ENHANCE LEARNING

Category of Submission: Short paper

Name & address: Dr. Tock Keng Lim  
National Institute of Education  
Nanyang Technological University  
469 Bukit Timah Road, Singapore 259756

Tel: 4605254  
Fax: 4687945  
E-mail: LIMTK@NIEVAX.NIE.AC.SG