No. 58 November 1986

A·C·E·R NEWSLETTER

Edited by Peter Jeffery and John King

Computerized Adaptive Testing

Assessment plays an important role in the education of students because it is the means by which a teacher obtains relevant information about the educational progress of his or her students. Information gained from assessments of students' learning and performance enables the teacher to make valid decisions concerning students' needs and directions for future learning. Many forms of assessment include the use of achievement tests that can be used to identify particular weaknesses in students' learning, and to identify students who are making progress and those who are not. On the basis of the resultant test information the teacher might decide to focus his or her teaching so that more assistance is directed towards particular students who are not doing well.

In the past two decades considerable progress has been made in the development of educational assessment methods that analyse test data in terms of item response theory (IRT), sometimes called latent trait theory. In IRT the interaction between a student and the items (questions) of a test are modelled statistically to provide a link between the student's responses to the items and the theoretical constructs (latent traits) assumed to underly the test items.

The IRT model of student–item interactions that ACER's project on *computerized adaptive testing* is concerned with is known as the Rasch model, which is the simplest of a number of IRT models. A student responding to the items in a test is assigned an ability parameter whose estimate locates the student on the underlying latent scale, and each item administered to the student is assigned a difficulty parameter whose estimate locates the item on the same underlying latent scale. That is, both persons and items are located on the one latent scale, and so valid comparisons can be made between performances of students and items. It is worth noting that in more complicated IRT models each item is characterized by parameters additional to item difficulty. However, these models do not have the attractive measurement properties of the Rasch model and are more difficult to apply to practical testing problems.

An important development in recent years has been computerized adaptive testing (CAT), which joins IRT methods with adaptive testing procedures. Moreover, the development of microcomputers and their introduction into schools has meant that CAT is now a possibility in classrooms. Adaptive testing with the Rasch model is a process of administering items that differ only in their difficulty to students on an individual basis, which capitalizes on a student's responses to previously administered items. One begins with a large item pool that is representative of an underlying latent trait or construct, then items are selected from the pool and administered to a student one at a time. After each item is administered, the student's response is scored as correct or wrong, the student's position on the latent trait scale is estimated, and a new item is selected from the pool to obtain a better estimate of the student's location in the scale. In this way, the items selected for administration to the student may be tailored to the student's level of proficiency so that the student is presented with only the number of items required to estimate the student's trait level or position on the latent trait scale to a predefined level of accuracy. Although different sets of items may be administered to different individuals, the resulting estimates of proficiency will always be represented on the one common latent scale.

Adaptive testing is well suited to implementation on microcomputers because it requires an extensive storage facility for the item pool, an efficient means for administering the items, a reasonably powerful computational facility to score and analyse the student responses, and a reporting facility that can produce tailored reports on students quickly and efficiently. Moreover, CAT provides a worthy facility for innovation in classroom practices: it could free the teacher from having to administer tests to the class as a whole at the same time—individual students could take the test when they were ready to do so; students could be provided with instant feedback by the computer about how well they were progressing in the course; the computer could automatically update information about students' performance, and this infor-

ACER Publications

ACER Psychological Catalogue 1986–1987.

ACER Special Education Bulletin 1986.

ACER Special Education, Health, and Community Services Catalogue.

Australian Education Index, Volume 29, Number 2, June 1986, and Number 3, September 1986.

PAT Reading Manual.

Priorities in Language Education: a Survey of Community Views. Kevin Piper & Hilary Miller. ACER Research Monograph No. 28.

School Organization and the Quality of Schooling: a Study of Victorian Government Secondary Schools. John Ainley, Rodney Reed, & Hilary Miller. ACER Research Monograph No. 29.

mation could be made readily available in the form of tailored reports to teachers and school.

This ACER project has three main aims: (1) the development of a comprehensive system of microcomputer programs in the BASIC language for CAT, initially for the Apple II family of computers, which are portable and easily used in a classroom environment; (2) the development of computerized adaptive tests based on the existing paper-and-pencil ACER mathematics tests—the AM Series and the Profile Series; and (3) examination of some research problems concerning the use of CAT.

The first aim is essentially a problem of design, and provides a CAT scheme of sufficient generality that contains utility programs for storage and retrieval of test items, programs for administering, scoring, analysing and reporting the results of CAT, and a teacher's manual that describes the rationale of the scheme and how to apply it. The programs will be provided on floppy diskettes which will be inserts in the teacher's manual.

The utility programs focus on the teacher's individual data management needs for assessment and improving instruction. The data base of test items is file oriented, which allows the creation of pools of items within the one computer environment. Moreover the CAT scheme is able to handle items with significant pictorial content, and to link items to form tests according to criteria specified by the teacher.

The computer programs for administering, scoring, analysing, and reporting the results of CAT are based on recent research findings. Although there have been many important developments in the research literature in this area, practical applications of this research are still limited and not widely implemented. Problems we are working on at present include the optimum entry level or starting point for administration of the first item, the item selection rule for succeeding items, and the termination criteria for ending a testing session. The project is examining these problems from both classical and Bayesian statistical perspectives.

The second aim of the project is the development of computerized adaptive test versions of two existing series of ACER mathematics tests. These tests are providing the item base for development and refinement of the CAT system. At the completion of the project, the two series of tests will be marketed by the ACER as alternative versions of the existing paper-and-pencil tests. Information on the use of the two versions should provide data for assessing advantages and disadvantages of the two modes of presentation and the likely acceptance by Australian teachers and students of computerized tests and computerized materials in general.

The third aim of the project is concerned with a number of research problems in the use of CAT. Specifically the project is focusing on the differences between paper-and-pencil testing and computerized testing, and the effects these differences have in the assessment of students' proficiency—available evidence suggests that there could be marked individual variation between the two conditions, and that these might influence validities more than other psychometric features such as the reliability of tests.

We are also studying the extent to which CAT improves test-taking performance (speed of initial response, time between responses, number of attempts at items, and time to complete testing sessions with individuals and groups), the extent to which computerization of tests may be expected to change the nature of a test, and the kinds of information required to assist in the interpretation of results obtained from CAT.

For further information regarding this project contact Mr George Morgan at ACER.

What is the Australian Education Index?

Q. How do you find out about 'it'? . . . 'Your pet topic'?

Q. How do you find out about (for example) . . .
Curriculum development in computer science education
Retention rates in secondary schools
Parent participation in reading teaching
Effect of classroom environment on academic achievement
Community involvement
Teaching methods
Environmental education
Multicultural education
Migrant education
Aboriginal education
Enrolment projections
Mathematics teaching and females
Teaching reading and writing
English as a second language
Teaching second language
Teaching science

A. Consult AEI!
The Australian Education Index covers everything from 'Ability grouping' to 'Zoning'. It is a comprehensive index to published and unpublished material relevant to Australian education

The index is published quarterly with an annual cumulation. It has been produced regularly by the ACER Library since 1957 and has been available for online searching through the Australian Information Network (AUSINET) since 1979. Material relevant to all levels and areas of Australian education is identified and indexed for retrieval by authors, institution, and subject.

Approximately 350 separate Australian periodical titles are scanned to identify relevant material for inclusion in the index. Books, reports, conference proceedings, and monograph series are also included. Material published by Australian authors or published overseas about Australia is also included.

Two specialist clearing houses contribute to the AEI. The Australian Clearing House on Library and Information Science identifies and indexes material relating to library and information science and the TAFE National Clearinghouse contributes indexed entries on vocational and career education.

Online User Courses are available (see details elsewhere in this newsletter).