

# Computerized Adaptive Testing

Assessment, diagnosis, and placement by computer may soon replace paper-and-pencil testing.



James I. Garrett

“**T**he test of the future ... TODAY!” A dozen years ago, a small cadre of University of Minnesota graduate students used that slogan to attract volunteers to the first experimental trials of computerized adaptive testing (CAT). Two years ago some of those same researchers participated in an American Educational Research Association symposium entitled “Computerized Adaptive Testing: The Future Is Now.” This year, test publishers will offer the first practical applications of adaptive testing for use in educational settings. Some advocates believe that within 15 years this form of testing will completely displace paper-and-pencil tests in many programs.

## Features and Benefits

The two features that distinguish computerized adaptive testing from its conventional counterpart are implied in its name: *computerized test administration* and *adaptive test design*. First, the examinee uses a standard keyboard or specially designed auxiliary device to answer questions that are displayed on the computer screen. Second, tests are individually adjusted to the ability or achievement level of the examinee. In effect, the computer program chooses a more difficult question after each correct answer and an easier question after each wrong answer.

The benefits of computer administered adaptive tests are both psychometric and practical. The psychomet-

ric benefits can be summed up in one word: efficiency. A well-constructed adaptive test attains a specified level of measurement precision in about half the length of time a conventional test would require to reach the same level. This is attributable to the adaptive feature; by tailoring the choice of questions to match the examinee's ability, the test bypasses most questions that are inappropriate in difficul-

ty level and contribute little to the accurate estimation of the test-taker's ability.

This psychometric efficiency can be used to (1) reduce the length of the test, retaining an acceptable level of measurement precision while cutting the average testing time in half; (2) leave testing time unchanged to allow greater precision in measurement; or (3) leave both testing time and measurement precision unaltered and extend the breadth of content of the testing program.

Of the numerous benefits of computerized test administration, two stand out: standardization and immediacy. Standardization means simply that computer-controlled testing has the potential to provide greater standardization of administrative conditions than paper-and-pencil tests normally do. Immediacy refers to the computer's inherent ability to eliminate time-consuming and error-prone intermediate stages between test administration and the availability of results. Immediately, and virtually without error, the computer can score tests, convert the scores into norm- or criterion-referenced interpretations, print the results, and record them in its memory.

### **The State of the Art**

There are five components essential to an adaptive testing program.

1. A *psychometric foundation* is a valid, defensible theoretical basis for administering different questions to different people yet expressing the results on a single scale. This foundation already exists in the body of item response theory as advanced by Lord (1970, 1980), Rasch (1960), and Wright and Douglas (1975). The single contribution of item response theory to computerized adaptive testing is that it provides a basis for locating both test questions and examinees on the same scale and, therefore, tailoring the difficulty of the tests to the ability level of the examinees.

2. An *item bank* is a large set of test questions that measures the domain of interest and has been calibrated using the methods of item response theory.

It is desirable for the number of items in the bank to substantially exceed—say, by a ratio of five or ten to one—the number of questions an individual examinee will encounter (Ree, 1977). The state of the art of this component depends on the testing program itself. In the educational measurement arena, several large-scale programs for assessing individual achievement have already assembled numerous test questions that have been calibrated using the methods of item response theory. Even though adaptive versions of these achievement tests have not been developed, it would seem that the item bank resources are readily available.

3. A *strategy* is a set of methods for successively approximating the examinee's ability and sequentially choosing the best test question to administer at every stage of the test. Strategies of adaptive testing have been the subject of research and development for some time (Lord, 1970; Weiss, 1974; Crichton, 1981). Although there is room for still more research in this area, enough is known about some strategies to have confidence in their use (Wetzel and McBride, 1983).

4. A *delivery system* is a matched set of computer hardware and software capable of presenting test questions to the examinee, performing the computations necessary to approximate the examinee's location on the measurement scale, selecting the best question to present next, determining when to stop testing, and recording and reporting the results in the medium and format necessary to satisfy the requirements of the application. Experimental computer systems for adaptive testing have been available for almost two decades, and several systems capable of operational use exist today (Assessment Systems Corporation, 1984; Quan and others, 1984).

5. An *experience base* is a body of research, development, and empirical evidence that justifies confidence in the usefulness and validity of adaptive testing. This base is fairly small, but it is growing rapidly along with the progress of major projects being developed by the Department of Defense

**“Some advocates believe that within 15 years computerized adaptive tests will completely displace paper-and-pencil tests in many testing programs.”**



Christine L. Roberts

and commercial test publishers. The psychometric efficiency of computerized adaptive testing has been demonstrated several times (McBride and Martin, 1983; Moreno and others, 1983; Urry, 1977), as has its validity as a measurement procedure (Moreno and others, 1983) and its usefulness for predicting performance (Vicino and Hardwicke, 1984).

In summary, the state of the art is sufficiently advanced to encourage pioneering applications of computerized adaptive testing to the measurement and assessment needs of the educational community.

### Potential Applications

Three broad classes of applications seem appropriate:

**Assessment Tests.** The Department of Defense is preparing a computerized adaptive version of the aptitude tests it uses for selecting enlisted personnel (see Vicino and Hardwicke, 1984). In a similar vein, the Psychological Corporation is developing an adaptive version of the Differential Aptitude Tests for vocational counseling in schools.

Survey achievement tests could be administered in a computerized adaptive format as well. The significant advantage would be the considerable savings of testing time, since achievement batteries often require four or more hours per pupil. Although this application is appealing, it is impractical at this time since it requires a much larger number of computer terminals than is typically available in schools. A related and logistically feasible application might be the use of adaptive "locator" tests to identify individual candidates for out-of-level achievement testing.

Basic skills or competency testing is mandatory in many jurisdictions and is usually conducted in a large-group setting. A computerized adaptive competency testing program could be conducted throughout the year on a walk-in or by-appointment basis, or in a manner that encourages students to take the test when they feel ready. Again, the advantage would be significant time savings per pupil, and the obstacle is the impracticality of testing large groups simultaneously.

**"Immediately, and virtually without error, the computer can score tests, convert the scores into norm- or criterion-referenced interpretations, print the results, and record them in its memory."**

**"... computers will not replace the teacher, but instead will serve as an electronic 'workbook' that is especially useful for daily assessment and diagnosis of pupils."**

*Diagnostic Tests.* An especially appealing potential application of adaptive testing is for individualized diagnostic testing of academic knowledge and skills to identify deficiencies and prescribe remediation. That can be done much more efficiently with a computer than in a paper-and-pencil format, and it might make the computer a valuable tool for the classroom teacher.

*Placement Testing.* At the secondary and postsecondary levels, computerized adaptive testing could be used to identify strengths and weaknesses in each new student's mastery of the prerequisites of multi-level or elective courses. The Educational Testing Service is pioneering this application with an adaptive battery of placement tests for use in community colleges (Ranbom, 1985). Similar systems may be equally useful to place transfer students and students who apply for admission to advanced courses.

### The Future of Computerized Testing

Large-scale applications of computerized testing are nonexistent today for a number of reasons, including the cost of computers themselves and the cost

of access to the fairly large storage capacity required to support computerized testing. Microcomputers are commonly found in American elementary and secondary schools, but they are usually located in computer laboratories rather than classrooms.

In the not-too-distant future, I expect significant changes in the availability of computers in the classroom. These changes will include evolutionary reductions in the cost of computers, revolutionary advances in the capacity of inexpensive disk storage, and increased use of local networks to allow communication and sharing of expensive resources among all the computers in a school or district.

As these changes occur, there will be a movement toward establishing electronic resource centers in each school or district, which will be accessible from very inexpensive terminals or microcomputers in each classroom. These resource centers will distribute instructional courseware and testing software for the electronic medium at a cost comparable to that of today's textbooks and resource materials.

That evolution will take place gradually. In its later stages, I expect that every classroom will have several computer terminals or microcomputers—perhaps as many as one per pupil. These computers will not replace the teacher, but instead will serve as an electronic "workbook" that is especially useful for daily assessment and diagnosis of pupils.

This prospect brings with it a challenge for educational planners to anticipate the future, identify the most promising testing uses of computers, and encourage the development of these tests of the future. □

### References

- Assessment Systems Corporation. *The MicroCAT Testing System*. St. Paul, Minn.: Assessment Systems Corporation, 1984.
- Crichton, L. I. "Effect of Error in Item Parameter Estimates on Adaptive Testing." Doctoral dissertation, University of Minnesota, Department of Psychology, 1981.
- Lord, F. M. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.
- Lord, F. M. "Some Test Theory for Tailored Testing." In *Computer-Assisted Instruction, Testing, and Guidance*. Edited by W. H. Holtzman. New York: Harper and Row, 1970.
- McBride, J. R., and Martin, J. T. "Reliability and Validity of Adaptive Ability Tests in a Military Setting." In *New Horizons in Testing*. Edited by D. J. Weiss. New York: Academic Press, 1983.
- Moreno, K.; Wetzel, C. D.; McBride, J. R.; and Weiss, D. J. *Relationship Between Corresponding Armed Services Vocational Aptitude Battery (ASVAB) and Computerized Adaptive Testing (CAT) Subtests* (NPRDC TR 83-27). San Diego: Navy Personnel Research and Development Center, 1983.
- Quan, B. L.; Park, T. A.; Sandahl, G.; and Wolfe, J. H. *Microcomputer Network for Computerized Adaptive Testing* (NPRDC TR 84-33). San Diego: Navy Personnel Research and Development Center, 1984.
- Ranbom, S. "E.T.S. Ready: Computer-Based 'Test of the Future.'" *Education Week*, May 22, 1985, p. 9.
- Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen and Lydiche, 1960.
- Ree, M. J. "The Effects of Item Calibration Sample Size and Item Pool Size on Adaptive Testing." *Applied Psychological Measurement* 5 (1977): 11-19.
- Urry, V. W. "Tailored Testing: A Successful Application of Latent Trait Theory." *Journal of Educational Measurement* 14 (1977): 181-186.
- Vicino, F. L., and Hardwicke, S. "An Evaluation of the Utility of Computerized Adaptive Testing." Paper presented at the convention of the American Educational Research Association, New Orleans, La., April 1984.
- Weiss, D. J. *Strategies of Adaptive Ability Measurement* (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974.
- Wetzel, C. D., and McBride, G. A. *Influence of Fallible Item Parameters on Test Information During Adaptive Testing* (NPRDC TR 83-15). San Diego: Navy Personnel Research and Development Center, 1983.
- Wright, B. D., and Douglas, G. A. *Best Test Design and Self-Tailored Testing* (Research Memorandum No. 19). Chicago: Statistical Laboratory, Department of Education, University of Chicago, 1975.

**James R. McBride** is Director, Computer Based Testing Group, Measurement Division, The Psychological Corporation, 1250 Sixth Avenue, San Diego, California 92101.

Copyright © 1985 by the Association for Supervision and Curriculum Development. All rights reserved.