COMPUTER ASSISTED TESTING: AN ORDERLY TRANSITION FROM THEORY TO PRACTICE

RICHARD H. MCKILLIP
VERN W. URRY
U.S. Civil Service Commission

The United States Civil Service Commission is responsible for examining applicants for Federal jobs throughout the world. It examines almost two million persons and makes about 200,000 placements annually.

The Commission's investment in computerized adaptive testing research and development is a significant one. This exciting and innovative program is currently budgeted at almost $200,000 per year. This expenditure comes at a time when Federal agencies' budgets are most austere and when resources are sorely needed to respond to the increasing challenges faced by conventional examining methods.

The Commission's investment in computerized adaptive testing is based primarily on the potential payoff in improved employee selection and placement. The large numbers of examinations and applicants makes computerized adaptive testing an economical, practical vehicle for improved measurement. The answer to attacks on tests in the employment situation is complex; the economic and social implications of this problem are enormous. Unquestionably, however, the greatest benefit both to the employer and to the employee lies in better measurement, not in less measurement. Every improvement in the selection and placement processes should contribute to the economic health of the employer, the psychological well being of the affected individual, and the welfare of society. Computer technology offers not only an opportunity to make significant improvements in employment decisions but also a better means of assessing the effects of such improvements.

While there are problems yet to be solved, computerized adaptive testing is well on the way to implementation.

As conventional approaches to test construction are modified in light of developments in latent trait theory, computerized adaptive testing becomes more and more feasible. The Rasch Model showed capabilities for computerized adaptive testing in the special case where all items discriminated equally and were unaffected by guessing. This special case was simply not practical to expect in available test items (Urry, 1970). Since item requirements for three-parameter logistic or normal ogive models can be met with existing items (Lord, 1970), computerized adaptive testing can be implemented. The implementation can be cost effective (i.e., the number of test items administered is substantially reduced vis-a-vis conventional testing) when certain rigorous item bank specifications can be met (Jensema, 1975). The determination that the item bank specifications can be met with existing items is contingent upon a new look at conventional item statistics and their relationship to model parameters. It has become apparent that the distortions caused by guessing result in severe underestimates, particularly of item discriminatory powers (Urry, 1975). Reliable estimates of parameters can now be made (Gugel, *et al*, 1975). An algorithm exists that will allow on-line computer-interactive item calibration (Schmidt & Urry, 1975).

Problems remain in tailoring test batteries to specific occupational requirements and in adequate coverage of job-related abilities. Of serious concern are the time and dollar respources that are needed for comprehensive measurement. The improved medium of presentation inherent in the hardware will facilitate resolution of these problems; for example, new item types and audio input possibilities.

Application of computerized adaptive testing in civil service examining has several desirable features.

Job relatedness. With multivariate test item banks, it is feasible to interpret scores on specific abilities in terms of differential occupational requirements. This then enables the employer to test a large number of abilities and to weight these abilities in accordance with their importance for success in specific jobs. The employer can array applicants across a large number of jobs and select in terms of priority, thus maximizing the utility of the selection process.

Standardized Examination Administration. Individual differences among administrators under conventional testing make error variance due to unstandardized administration largely unavoidable. Since administration procedures can be programmed under individualized testing, standard conditions can be better maintained.

Compromise of Examination Materials. Under computerized adaptive testing, examination questions are located in a central computer. No test booklets are used, therefore none can be taken from the examination room. As a result, the security of tests and test questions can be maintained more easily. Different individuals will receive different sequences of items, reducing the likelihood of cheating.

Improved Administrative Procedures. Test booklet printing, storage, and distribution costs become inconsequential.

Examination Scheduling. Tests can be administered on a walk-in basis since different tests can be administered simultaneously. The shortened testing time makes possible the administration of a multiple abilities battery in the time now required to examine for a single ability. Further, if selection is specific to a given position, individualized testing for the required abilities can be accomplished in a manner that minimizes the time of testing while maximizing the job relatedness of a final weighted score.

Power Conditions of Examination. Tests of ability should be power tests. However, due to administrative considerations, i.e., scheduling, space restrictions, etc., conventional tests of ability are usually speeded to a certain degree. Under computerized adaptive testing, the power conditions required by this type of test can be ensured.

Test-Taking Motivation. Test-taking motivation and, consequently, test performance may be impaired when the level of difficulty of the examination material is inappropriate to the level of ability of the examinee. In conventional testing, the examination is constructed for an entire population. This method of construction necessarily leads to inappropriate question difficulties when a conventional test is presented to a given examinee. In computerized adaptive testing, the difficulty level of the questions is matched to the level of ability of the examinee.

Improving Examinations. The current conventional testing technology is the product of more than fifty years of research and development. Substantial improvements have been less frequent with the passage of time. This calls for a rather dramatic change in testing procedure. At present, the appropriate change would be towards an individualized testing technology. Certainly greater experimental control and a thorough monitoring of the measurement process is made possible through the aid of this new medium.

Improving Personnel Decisions. When a computer interactive network has been established for individualized testing, one has necessarily established a vast data accession network to effect immediate evaluation of the personnel decision making process. Optimization in the decision-making process is the natural extension of events when many sources of information are available to a central computer and are readily accessible for analysis by the personnel researcher and personnel specialist.

It appears, at this time, that computerized adaptive testing research has progressed to the point where implementation will be feasible. In Fiscal Year 1976, a comprehensive cost analysis will be undertaken. Preliminary estimates are favorable. For example, computer connect time in testing in one ability area now costs less than forty cents per examinee. It is reasonable to expect that cost to drop as the program progresses. Current plans call for fully operational computerized adaptive testing by 1980. At that time, it is expected that the examination for most entry-level professional and administrative jobs will include a test battery administered in the computerized adaptive system. Approximately 200,000 applicants currently file for these jobs. It will take until 1980 to get ready for an examination of this scope and number of participants.

My colleagues this morning will address some of the progress we have made in solving technical problems associated with the program.

REFERENCES

Gugel, J. F., Schmidt, F. L. & Urry, V. W. *Effectiveness of the ancillary estimation procedure.* Paper presented at the conference on Computerized Adaptive Testing, Washington, D.C.: June 1975.

Jensema, C. J. *Bayesian tailored testing and the influence on item bank characteristics.* Paper presented at the conference on Computerized Adaptive Testing, Washington, D.C.: June 1975.

Lord, F. M. Item characteristic curves estimated without knowledge of their mathematical form – a confrontation of Birnbaum's logistic model. *Psychometrika,* 1970, 35, 43-50.

Schmidt, F. L. & Urry, V. W. *Item parameterization procedures for the future.* Paper presented at the conference on Computerized Adaptive Testing, Washington, D.C.: June 1975.

Urry, V. W. A Monte Carlo investigation of logistic mental test models. (Doctoral dissertation, Purdue University, 1970). *Dissertation Abstracts Internatinal,* 1971, 31, 6319B. (University Microfilms No. 71-9475)

Urry, V. W. Approximations to item parameters of mental test models and their uses. *Educational and Psychological Measurement,* 1974, 34, 253-269.

Urry, V. W. *The effects of guessing on parameters of item discriminatory power.* TN-75-2 Washington, D.C.: U.S. Civil Service Commission, Personnel Research and Development Center, 1975.