JATT Special Issue on Adaptive Testing: Welcome and Overview

Nathan A. Thompson

Assessment Systems Corporation

Computerized Adaptive Testing (CAT: van der Linden, 2010) is a sophisticated methodology for designing, delivering, and scoring assessments. It has nearly four decades of research behind it (e.g., Betz and Weiss, 1973), and yet remains underutilized, having only been applied more recently to certain industries such as pre-employment testing and healthcare. However, interest in CAT is growing, especially in the K-12 educational assessment industry in response to calls for integration of technology and efficiency. In light of this expansion, the *Journal of Applied Testing Technology* has organized a special issue on CAT with a focus on practical applications and concerns. As guest editor, I appreciate the opportunity to share the articles in this special issue with readers following in line of the 2010 special issue that focused on assessment of students with disabilities.

CAT operates by leveraging the programmed interactive nature and computational powers of computers to make tests more efficient. The technology of CAT provides a number of benefits, including a better examinee experience and motivation, but the most noted benefit is that CATs typically require only half as many items as conventional fixed-form (i.e., linear) tests to obtain the same level of score precision (Weiss & Kingsbury, 1984). As computers become more and more widespread, item response theory gains wider use, and the Internet allows an efficient way of both delivering CATs and recording their results, CAT becomes more and more feasible and useful to a greater range of organizations.

However, CAT should not be viewed as a panacea because the methodology will not be appropriate for all programs. Items must be parameterized with item response theory (IRT), which necessitates pretesting of items, often with at least several hundred examinees. Also, because items must be scored in real time, subjectively scored items are often not feasible. Additionally, innovative items with complex non-IRT scoring are often a poor fit.

This special issue presents four articles that discuss the development and application of existing CAT testing programs and the issues they have faced. They present a range of constructs as well as an international perspective. In addition to K-12 educational assessment in the United States, the organizations include an internationally recognized higher education admission test, an educational formative assessment in Denmark, and a workforce development test in Singapore.

Kingsbury and Wise (this issue) describe research conducted for an organization that has been using CAT for nearly 30 years. In their study, they raise the question of score scale stability, that is, are scores reported on a consistent scale over long periods of time? The results suggest that for this program this was the case.

Rudner and Guo (this issue) examine measurement decision theory as an alternative model to the ubiquitous item response theory CAT foundation. This model has potential utility for situations faced by programs with practical constraints such as low sample sizes, but that still require an efficient way to classify examinees into groups along an underlying continuum.

Wandall (this issue) describes the application of CAT to a national assessment program with a formative intent, including discussions of practical issues such as software platform and examinee score reporting. This article may be of great interest for readers working in primary and secondary educational assessment, especially with regards to some of the Race to the Top initiatives in the United States. However, many of the issues Wandall discusses are relevant to all organizations utilizing CAT.

Jacobsen et al. (this issue) describe the development of a CAT system for assessing basic employability skills in the Singapore workforce. The comprehensive overview provided of the development process is informative for organizations developing or planning to develop a CAT. They

discuss important validity issues outside the CAT algorithm, such as standard setting, fairness and sensitivity review, and score reporting.

I would like to personally thank the authors for the effort they have put into these articles. They each provide a practical resource for testing organizations considering or developing CAT, and therefore contribute to the foundation that is being laid in the test publishing field for even more widespread application of this important technology for making tests more efficient and making the most out of the technology available in computerized test administration.

References

Jacobsen, J. Ackermann, R., Egüez, J. Ganguli, D., Rickard, P., & Taylor, L. (2011). Design of a Computer-Adaptive Test to Measure English Literacy and Numeracy in the Singapore Workforce: Considerations, Benefits, and Implications. *Journal of Applied Testing Technology*, this issue.

Kingsbury, G.G., & Wise, S.L. (2011). Creating a K-12 Adaptive Test: Examining the Stability of Item Parameter Estimates and Measurement Scales. *Journal of Applied Testing Technology*, this issue.

Rudner, L.M., & Guo, F. (2011). Computer adaptive testing for small scale programs and instructional systems. *Journal of Applied Testing Technology*, this issue.

Wandall, J. (2011). National tests in Denmark – CAT as a pedagogic tool. *Journal of Applied Testing Technology*, this issue.

Weiss, D.J. & Betz, N.E. (1973). *Ability measurement: Conventional or adaptive?* (Research Report 73-1). Minneapolis, MN: University of Minnesota, Psychometric Methods Program.

Weiss, D.J. & Kingsbury, G.G. (1984). Applications of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*, 361-375.

van der Linden, W.J. & Glas, C.A.W. (2010). *Elements of Adaptive Testing*. New York, NY: Springer.