Quality Control and Test Security in Standardized Assessments

With Multistage Test Designs

Alina A. von Davier

Educational Testing Service
Princeton, USA – avondavier@ets.org

A multistage test (MST) is a computer-based assessment that may be thought of as a compromise between a linear test and a computer-adaptive test (CAT). As such, MSTs may be vulnerable to at least some of the major security threats associated with each of these types of test (e.g., copying for linear tests and item pre-knowledge for CATs). The degree of vulnerability of any particular MST to these threats, as well as others, will depend (among other things) on details of the MST assembly and administration design. To supplement these preventative measures, routine statistical monitoring of response and timing data for items, modules and tests, as well as the screening of performance of individual test takers and clusters of test takers, is essential.

For monitoring and screening procedures to be effective, it is critical that statistical baselines be developed for the different characteristics of the data that are being monitored. Since an MST has a different structure than either a linear test or a CAT, it is reasonable to expect that test takers' behavior for MSTs will differ systematically from behavior observed for either linear tests or for CATs. In this presentation I discuss two concrete analyzes, one that identifies people who are outliers and one that identifies items that are outliers. For the first approach we use the scores from the variable section of the test (and make use of the MST design) to produce an expected score. For the second approach I present the use of traditional quality control tools such as CUSUM charts to identify items that have been exposed. What we are looking for are data patterns that deviate substantially from our baselines in ways that are consistent with what would be expected from behavior that may lead to invalid scores.

Keywords: Multistage Adaptive Tests, CUSUM charts, Prediction