Innovations in Testing at Other Credentialing Organizations
Panel Summary Report

Panelists:

Lawrence Rudner, PhD
Vice President of Research and Development (past – see Speaker page for current position)
Graduate Management Admissions Council

Ada Woo, PhD
Director of Measurement and Testing
National Council of State Boards of Nursing

Developing and implementing innovative testing solutions for an ongoing high-stakes testing program involves some unique challenges outside of test development, including business planning, budgeting, staffing, and candidate communications. During this session, assessment experts from the Graduate Management Admissions Council (GMAC) and National Council of State Boards of Nursing (NCSBN) shared their experiences introducing innovations into their existing high-stakes testing programs.

GMAT Innovative Items Lessons Learned

Dr. Lawrence Rudner began by providing some background information on the organization. GMAC is an international non-profit organization whose flagship examination, the Graduate Management Admission Test (GMAT), is the most widely used standardized assessment for graduate business administration programs. Over 250,000 test-takers take the GMAT each year.

The test historically offered assessment in three areas: analytical writing, quantitative reasoning, and verbal reasoning. In 2012, integrated reasoning was added to address the increasingly data-driven world that now requires managers to make decisions and judgments while combining verbal and quantitative reasoning to solve problems.

To determine the skills to be tested in this new section, over 700 management schools were surveyed. As a matter of practice, GMAC conducts surveys of its members every 5 years to ensure that the exam remains relevant. It was from these initial surveys that they discovered the need to begin working on the integrated reasoning portion of the examination.

In designing the new section, GMAC began with 9 possible item types. However, after testing out the items, they decided on the following four new item types:

- Data table interpretation
- Graphics interpretation
- Multi-source reasoning
- Two-part analysis

Dr. Rudner observed that they succeeded in engaging their audience by surveying them to determine what needed to be tested, field-testing items, and developing a scalable model. However, they were less successful in other areas such as scoring the new integrated reasoning section as part of the total rather than as a separate section, using dichotomous scoring rather than awarding partial credit on multiple true-false items, and failing to take advantage of assessment engineering. He also stressed that because of the monopoly mindset, GMAT did not do a return-on-investment analysis and planned to do this regardless of the cost. He recommended that testing organizations think about the entire process first and clarify their goals and objectives.

NCSBN Lessons Learned

Dr. Ada Woo began by providing some background information on the organization, the members of which are the various state boards of nursing. The NCSBN's mission is public protection. The NCLEX is the NCSBN's flagship assessment with almost 300,000 exams administered each year.

Dr. Woo provided a timeline of the NCSBN's current efforts to develop innovative items that began in 2011. The rationale for exploring new item types was to help nurses develop clinical judgment and evaluate critical thinking skills that current testing methods address only indirectly.

In developing a classification framework for prototypes of new items, the NCSBN began by identifying specific clinical judgment skills and mapping them to various item types (e.g., multiple-choice, multiple-response, drag-and-drop, hot-spot, audio, graphical, video). Dr. Woo shared a matrix that illustrated which item types best measured each of the desired skills. She also showed a flowchart of their current item development process: 1) prototype development; 2) usability and pilot testing; 3) developing scoring models; and 4) acceptance or rejection of items types. She noted that the NCSBN is currently in the prototype development phase of this process.

Dr. Woo's concluding points emphasized the need to support all decisions with research and ensure that new item types measure the constructs being tested. It is also important to evaluate what staff resources are required and available. She stressed the need to collaborate with the test delivery vendor to leverage new technologies. She encouraged the involvement of policy-makers early in the process and the need for buy-in from all stakeholders. Finally, she recommended planning ahead and starting early because development of new item types is a 5- to 7-year process.

During the subsequent question-and-answer period, Dr. Woo acknowledged that the NCSBN did not fully realize the increased volume of work required and emphasized that a return-on-investment analysis is highly desirable. It is a mistake to engage in such efforts regardless of cost even though the test sponsor may experience an increase in good will as part of the equation. She also noted that a failure to incorporate results from new item types into the total score was a disincentive for programs to use the new items. She exhorted conference attendees to think through the entire process before they begin investing time and resources.