Message from the Associate Executive Director
Scoring the Certifying and MOC Exams

J. Anthony Parker, M.D., Ph.D.
Associate Executive Director

Psychometrics is the field of science that deals with evaluation of test results. The ABNM employs a psychometric firm to assist with evaluation of the certifying and maintenance of certification exams. The firm uses a mathematical method, known by its inventor's name, Rasch. The Rasch statistical model is widely accepted meaning that the results of our exam are considered credible. Credible evaluation is one reason the ABNM certificate is accepted as an indication of professional expertise by our colleagues, our patients, the payors, and government organizations like state medical boards.

The psychometric firm supplies us with information about both candidates and the questions. Questions that almost everyone answers correctly or incorrectly provide little information about candidate ability. Mid scoring question provide the most information about candidate ability. Some questions turn out to test capabilities unrelated to clinical nuclear medicine, especially questions with a complicated format where the ability to understand the question is more important than knowing the nuclear medicine concept being tested. Questions that are not in the mid range and questions that do not correlate with ability add noise to the test result and need to be eliminated from scoring.

For large test groups like the SAT or ACT it is practical to pre-test question. A question that has not previously been tested can be pretested on subset of exam takers. Since the exam group is so large, using only a tiny fraction of the test takers allows for very accurate statistics to be developed. A large number of questions can be pretested with each test taker only having one or a few test questions on his/her exam. Using pretested questions allows exam results to be provided shortly after the test.

For a small group like the ABNM that is not possible. Therefore, new questions are mixed with questions that have been previously tested. In the first part of the process, the psychometric firm performs a tentative evaluation providing statistics for each question. The ABNM Exam Committee then reviews these results with the aid of a psychometrician. Question that are poorly correlated with ability or which are too hard or too easy are discarded. The statistics often provide a good clue how to rewrite a question so that it will have better statistics. Comments about the questions provided by diplomates also assist in this process. Generally about 10% of the questions on an ABNM exam are discarded. Using the improved exam, the psychometric firm then provides the final scores for the test candidates. Candidates often ask why it takes so long to receive their results. The answer not only relates to having to calculate the results of the test twice, but also that the Exam Committee with the psychometrician needs to review the entire test to decide which if any questions should be discarded.

The new MOC evaluation, CertLink™, is going to be much different from the existing MOC exam (see ABNM: MOC Assessment). It will be a continuous process with a diplomate dashboard showing how the diplomate is doing. The intention is that CertLink becomes mostly a leaning process, deemphasizing the testing. A challenge for the ABNM and for the other boards that plan similar MOC assessment will be to make sure that the MOC evaluation remains credible to our colleagues, our patients, the payors, and government organizations. Demonstration of credible evaluation will be a key component of the pilot period. The ABNM along with the other CertLink boards believe this challenge can be accomplished, but it is likely to require innovations in the technology for assessment while learning.