**Adaptive Testing
Summary Report**

Presenter: Lawrence Rudner, PhD, MBA
Executive Vice President of Research
LMP Associates

Background

Computer Adaptive Testing (CAT) is a form of test delivery that utilizes computer technology to maximize the efficiency of an examination. An examinee's response to each question is analyzed immediately and used to select the next question they receive. The purpose is to maximize the precision of the exam, based on what is known about the examinee from prior questions. CAT affords the opportunity for a shorter exam which can provide more precise measurement than a fixed form test. Examinees who are clearly knowledgeable or clearly not knowledgeable will have shorter exam times, as the algorithm will be able to quickly determine their ability level. However, for those whose ability level is near the passing mark, longer test periods will be needed.

Key Points from Presentation

Dr. Lawrence Rudner began by presenting the algorithm design for computer adaptive testing. The examinee is presented a question of average difficulty. If the question is answered correctly, the computer will select a question of harder difficulty; if the question is answered incorrectly, the next question presented will be easier. The computer repeatedly returns to the item bank to select the next question. Correct responses to more difficult questions yield a higher score. This results in items being presented within a specific range targeted to examinee ability, which yields more precise measurement. The exam may also be designed as a multistage test, where sets of items rather than individual items are used. Performance on the set of items then determines the next set of items the candidate receives. Dr. Rudner noted that the advantages of CAT included a shorter, more precise exam; reduced seat time & exam cost; and enhanced test security because no two tests will be the same. Disadvantages include the need for a larger item bank which span a large difficulty range and that all items must be pretested (have statistics) for the algorithm to know what to select next.

Key Points from Breakout Session

The group reviewed how the selection algorithm works. Dr. Rudner confirmed that if the examinee starts off poorly because, for example, they are nervous, they can still "recover," because self-correction occurs during the exam. A potential use for this delivery system is in clinical vignettes, where the examinee is asked a series of increasingly difficult questions during management of a patient, similar to an oral exam. CAT can push the examinee with regard to their fund of knowledge. The group discussed advantages such as reduced seat time and cost, particularly for the General Pediatrics initial certifying exams which is 9 hours long. CAT also provides better test security, as no two examinees will receive the same set of questions. Disadvantages include having to enlarge our item pool, which would be a challenge for some of the subboards which have relatively small item pools. Another potential disadvantage is the examinee experience. Because the questions selected are targeted to the examinee's ability level, they will receive

a more challenging exam than they may be accustomed to. Also, examinees are not able to return to or review previously answered questions.

Conclusions

Overall the reaction toward CAT was positive and there was interest in exploring transitioning the initial certifying and MOC exams to this format in the future in order to increase measurement precision and reduce seat time and cost for examinees. The National Council of State Boards of Nursing (NCSBN) has utilized CAT for many years for the National Council Licensure Examination (NCLEX), and their representative, Dr. Ada Woo, Director of Measurement and Testing at NCSBN, indicated that it is crucial to educate examinees on how the examination works in order to make a successful transition to this testing model. Examinees need to understand their examination times will vary and that a longer examination time will not have yielded a different pass/fail result. Those who fail the examination may challenge the result, believing that if they had a few more questions, they could have demonstrated competence at the passing level. Finally, because the success of CAT is heavily dependent on the statistics of the items selected by the algorithm, item calibration is critical. To ensure appropriate statistics and item calibration, large sample sizes are needed. Therefore, CAT is likely feasible for only General Pediatrics, either initial certifying or MOC Part 3.