Auto Item Generation
Summary Report

Presenter: André De Champlain, PhD
Director of Psychometrics and Assessment Services
Medical Council of Canada

Background

Multiple choice questions are used extensively to measure medical knowledge, but are time consuming and expensive to create. Some newer testing methods, such as computer adaptive testing, require large item banks and extensive multiple choice item development. Traditional item development, where each question is individually written, reviewed, and revised by content experts, is expensive and time consuming. Auto Item Generation (AIG) is a process that utilizes a computer algorithm to generate items from cognitive models. Context experts are trained to develop these models rather than individual items.

Key Points from Presentation

Dr. André De Champlain provided an overview of the AIG methodology, which uses a three-stage process to generate multiple choice questions. First, the content necessary to generate the items must be identified and structured into a cognitive model. This can then be used to develop an item model, including a stem, the variables to be manipulated, and the answer options. Finally, a computer program can be used to generate a large number of items, check the constraints for each item, and remove illogical combinations. He noted that the Medical Council of Canada (MCC) has been using AIG for over 5 years. To date, over 10,000 items have been generated using more than 50 cognitive models. Pilot studies have shown that experts are not able to differentiate items that have been generated using AIG versus traditional methods. In addition, the MCC has found that on average AIG items are more difficult and have more discrimination power than traditionally written items. He noted that they are continuing to research, and they plan to develop several more applications to further automate the process to complete their transition. He noted preliminary findings indicate that AIG items are a few hundred dollars cheaper to develop than traditional items, but that the MCC has not yet completed a full cost-benefit analysis of AIG versus traditionally written items.

Key Points from Breakout Session

The process of building the cognitive and item models may result in improved item writing and more difficult and discriminating questions. Some in the group shared that this may be an excellent way to write questions around clinical guidelines, which are often slow to be adopted in clinical practice, and could quickly regenerate questions as guidelines are modified. Since the computer program can generate so many questions from one item model, the cost of individual items is lower and security concerns around recall of individual item details is diminished. It was noted that there are some subject areas, such as ethics, where building item models is more difficult. However, these same areas are often challenging using traditional item writing methods. Parameters must be carefully set to avoid items that too closely resemble each other. Care must also be taken to provide plausible distractors. AIG also requires a change in the role of subject matter experts from writing items to building cognitive and item models. This usually requires two training sessions, each lasting 2-3 days.

Conclusions

Overall the reaction toward AIG was positive. Both the breakout discussion group and conference attendees felt if the ABP chooses testing methods requiring an expansion of the current item pools, it would be worthwhile to investigate AIG.