TLEF Proposal: Building capacity for best-practices in assessment in the health sciences

1. Key Words: learning assessment, capacity building, item generation, collaboration

2. Abstract:

Purpose: We propose a novel grassroots project that involves collaboration between health sciences faculty and experts in assessment to efficiently generate test items. We will use a collaborative approach that creates an item building framework linked to course and curriculum objectives and a shared assessment resource.

Innovation: The item bank we will produce provides exciting opportunities to facilitate assessment of inter-disciplinary competencies. It also facilitates an efficient process for ensuring alignment between assessment, content delivery and curricular objectives. This will set up a structure to explore other methods of assessment in a collaborative, cross-faculty framework.

Impact on teaching and learning: The proposed approach guiding item generation also guides test interpretation. Test scores derived from this process should be more interpretable, allowing faculty to make better judgements about student performance and therefore providing clearer and more directed feedback.

3. Project Description:

Background

Assessment is essential to support and develop learners, yet students and faculty identify assessment as an area of concern (Nicol, 2010). "Assessment provides an indication of what the institution gives priority to in making judgements, it provides an agenda more persuasive than a syllabus or course outline and it therefore has a powerful backwash effect on all teaching and learning activities" (Boud, 2007, pg. 21). Robust assessment methods provide meaningful information for students, assisting them to develop required competencies, and also assist programs in evaluating the effectiveness of their curricula (Holmboe, 2010).

Assessment takes many forms, and includes thoughtful construction of multiple-choice questions (MCQs) that adequately assess the skills and knowledge of students. Testing clinical reasoning is a particular challenge as items that test these complex cognitive processes require a rich context and are often difficult to write (Epstein, 2007). Addressing this challenge is even more difficult for the majority of faculty who typically have little training in test development (Abdulghani et al., 2015). Current assessment practices in the health sciences are sub-optimal. Faculty tend to develop tests in isolation and in a highly inefficient manner leading to errors that can decrease the validity of a test (Abdulghani et al 2015; Verhoeven et al., 1999). The current University of Alberta Policy on Assessment and Grading outlines a number of key principles to guide faculty in assessment planning and development (UAPPOL, 2012). These policies, while designed to promote fair and high quality assessment, do not provide faculty with detailed information on how they are best achieved and as a result do little to support their implementation. While resources exist to assist individual faculty members in improving test development (e.g. Centre for Teaching and Learning, or Centre for Research in Applied Measurement and Evaluation), increasing this capacity within departments and faculties is more likely to meet the unique assessment needs of these educational units (Steinert, 2000).

To begin to address the challenges faced in providing efficient and high quality assessment, we propose a novel grassroots project that involves collaboration between health sciences faculty and experts in assessment to generate test items based on cognitive modeling (described below). We refer to this as the Interdisciplinary Educational Assessment of Learning (IDEAL) approach. With this approach, we will move away from test development in isolation toward a collaborative approach that creates an item building framework tagged to course and curriculum objectives and a shared assessment resource. This framework will clearly characterize the relationship between MCQ stems and response alternatives that can be used to improve the efficiency and quantity of quality items generated. These items may be utilized and adapted by faculty across multiple disciplines.

Theoretical framework

Item development begins when content specialists use design principles and guidelines to identify the knowledge and skills required to solve problems in a specific context. One way to structure these knowledge and skills is with a cognitive model. Leighton and Gierl (2007) defined the term 'cognitive model' in educational and psychological testing as "a simplified description of human problem solving on standardized educational tasks, which helps to characterize the knowledge and skills of students at different levels of learning" (p. 6). This model organizes the cognitive- and content-specific information so test developers have a framework for how students think about and solve tasks on tests.

Cognitive models can also be used to generate test items. Gierl, Lai, and Turner (2012) introduced the concept of a cognitive model for item development. This model helps link cognitive principles to item generation practices for the purpose of producing test items and is intended to highlight the knowledge, skills, competencies, and content required to create test items. To identify the content in the cognitive model, test developers are required to establish a framework that describes the knowledge, content, and reasoning skills required to create a problem-solving task (i.e., a test item). The knowledge and skills defined for the cognitive model are identified in an inductive manner by asking the developers to review a parent item and then to identify and describe information that could be used to create new items.

Three types of information required to solve the parent item are described. They include the problem and associated scenarios, sources of information, and features. These are specified as separate panels in the cognitive model (see Table for illustration). The top panel identifies the problem and its associated scenarios. The test developer begins by identifying the **problem and scenarios** specific to the parent item. The middle panel specifies the relevant **sources of information** needed to solve the problem. The bottom panel highlights the salient features, which include the elements and constraints. **Elements** contain content specific to each feature that can be manipulated for item generation. **Constraints** serve as restrictions that must be applied during the assembly task to ensure that content in the elements are combined in a meaningful way so useful items can be generated. By varying the sources of information and features within a problem, a large number of items can be generated based on the problem. The purpose of a cognitive model for item development is to organize the cognitive- and content-specific information in a structured representation of a task that, in turn, can be used to make inferences about how students are likely to think about and solve problems.

Table. Example from Medicine of creating an item building framework using a cognitive model, adapted from Gierl et al. (2012).

Problem and Scenarios	Postoperative fever with possible causes from urinary tract infection (UTI), atelectasis (A), wound infection (WI), pneumonia (P), deep vein thrombosis (DVT), deep space infection (DSI)
Sources of information	Patient demographics, physical examination, timing of fever, type of surgery
Features (e.g. timing of fever)	Element: 1-6 days Constraints: A=1-2 days, UTI = 2-3 days, WI, P, DVT = 2-3 days but may also occur 1-6 days, DSI = 4-6 days

Objectives

The long-term goal of this work is to strengthen and sustain best-practice in assessment by creating the IDEAL approach for the valid assessment of learning in health science faculties. This proposal focuses on MCQ item development that assess clinical reasoning by engaging a core group of faculty representing five Departments and Faculties at the University of Alberta. The objectives include:

1. Adapt the framework described above for MCQ generation to health science professional training programs.
2. Generate a large bank of MCQs that are potentially relevant to multiple disciplines within the health sciences that link to course and curricular outcomes within and across programs.

Methods

Project participants: Eight faculty representing Communication Sciences & Disorders, Occupational Therapy, Physical Therapy, and Pharmacy as well as experts in assessment from Education and Pharmacy.

Procedure: Experts in assessment (MG and KC) will provide instruction on the development of the cognitive modelling framework for generating MCQ items. Participants will use a nominal group technique (NGT) to generate the problems and scenarios to be solved, the relevant sources of information and the salient features as illustrated in the Table above (both within and across disciplines). NGT is a qualitative strategy to build consensus in small groups and has been used to develop educational programs and identify competencies in medical education. (Jones & Hunter, 1995; Fournier, 2014) Health science faculty will provide the frameworks to a trained graduate student assistant who will generate the items, tag each item to course and curriculum objectives and competencies, and create the item bank that is securely accessible to faculty across all health sciences disciplines. Finally, a trained graduate student will facilitate the production of an online resource detailing the item generation process and how to access and contribute to the MCQ item bank.

Outcomes/Deliverables

1. Hundreds of MCQs, each tagged to course and/or curriculum objectives, will be developed for each faculty member participating in the project.
2. Each item will be uploaded into a web-accessible and secure item bank.
3. A webinar will be produced, with open access to U of A faculty, detailing the item generation procedure and how to access and contribute to the item bank.
4. A Google group will be created to alert interested faculty about updates to the item bank as well as subsequent projects aimed at further enhancing assessment practices.

Evaluation

<u>Assessment of test/item quality</u>: The quality of summative tests and the MCQ items used to construct them will be established using participating faculty member's recent exams and be assessed by a trained graduate student. At the test level, alignment to course objectives and test reliability will be measured pre- and post- implementation of the project. Alignment to course objectives will be measured by comparing expected and observed tables of test specifications. Expected tables of test specifications use proportions to describe how course learning objectives are intended to be represented. Observed tables of test specifications use proportions to describe how tests actually represent course learning objectives. Comparing the expected proportions to the observed proportions facilitates calculation of the degree of alignment. Test reliability will be measured using Cronbach's alpha. At the item level, indexes of difficulty and discrimination will be compared pre- and post- project implementation. The distribution of item difficulties pre- and post- project implementation can be compared to determine how well they represent the range of item difficulties. Item discrimination indexes indicate how well items differentiate between those who have the knowledge needed to answer the question and those who don't. These indexes are like correlations in that they range between -1 and +1. Discrimination indexes should be positive. The distribution of discrimination indexes will be compared pre- and post- project implementation.

<u>Assessment of the perceived utility of the resources</u>: Faculty from each of the health science faculties will be recruited to complete the webinar and the procedure to access and contribute to

the item bank. They will complete a short survey pertaining to the clarity and utility of the webinar and the accessibility of the item bank. In addition, faculty will be asked the extent to which they intend to access and/or contribute to the item bank. The webinar will be adapted, based on this feedback.

Innovation, Collaboration and Sustainability

This collaborative process will build capacity within faculties to efficiently implement best-practices in assessment. This approach will complement existing on-campus resources for faculty to enhance their teaching. The outlined approach will increase the efficiency of item generation, improve the ease of linking items to competencies and promote cross-disciplinary collaboration and item sharing. It facilitates an efficient process for ensuring alignment between assessment, content delivery and curricular objectives. The item bank will be populated with items testing clinical reasoning that are relevant to many disciplines across campus. This will provide exciting opportunities to facilitate assessment of inter-disciplinary competencies. Finally, it will set up a structure to explore other methods of assessment in a collaborative, cross-faculty framework.

The Faculty of Rehabilitation Medicine (FRM) Teaching Interest Group (TIG) is a grassroots group of faculty that meet regularly to discuss and work on small projects that support faculty from various health science disciplines to enhance teaching and learning. Our group has collaborated with a variety of experts in teaching and learning and this project is an example of our successful cross-campus collaborations. Appended to this application are letters of support from the three Department Chairs within the FRM. Importantly, we are confident that this project will provide an opportunity for other health science faculty to either collaborate with the FRM TIG or implement this process within their own teaching groups.

There are two complementary components to our sustainability plan: (1) maintaining this collaborative practice of using cognitive models to develop frameworks for item generation and (2) growing the item bank. By generating items using a cognitive model that focuses the process on assessment principles and content (instead of content alone), faculty will be able to adapt this approach to their needs and support others within their teaching units. This process also facilitates course and curriculum mapping which will serve as an ongoing impetus to efficiently generate and tag items. Using a webinar to disseminate our process will encourage more faculty to implement this procedure and the item bank will continually grow. The existing collaborations within the FRM TIG will be used to disseminate this process and grow our group. Ongoing communication will occur through the FRM TIG Google Group listserv.

Impact on students

Student learning is enhanced with higher quality assessment and the process of using collaboratively developed cognitive models to guide item generation will increase the likelihood that tests measure intended learning objectives in a reliable, informative and efficient way. This is because MCQs linked to a cognitive model contain information in the response alternatives that can be used to diagnose specific errors in reasoning. This procedure saves time and enables faculty to use this information to generate more useful feedback as well as to guide future instruction. Ultimately, by engaging in the IDEAL approach to develop assessments, student learning should improve because faculty will be able to make better judgements about student performance and therefore provide clearer and more directed feedback.

References

Abdulghani, H. M., Ahmad, F., Irshad, M., Khalil, M. S., Al-Shaikh, G. K., Syed, S., ... & Haque, S. (2015). Faculty development programs improve the quality of Multiple Choice Questions items' writing. *Scientific reports*, *5*.

Epstein, R. M. (2007). Assessment in medical education. *The New England Journal of Medicine, 356*, 387-396.

Fournier, J. P., Escourrou, B., Dupouy, J., Bismuth, M., Birebent, J., Simmons, R., ... & Oustric, S. (2014). Identifying competencies required for medication prescribing for general practice residents: a nominal group technique study. *BMC family practice*, *15*(1), 139.

Gierl, M. J., Lai, H., & Turner, S. (2012). Using automatic item generation to create multiple-choice items for assessments in medical education. *Medical Education*, *46,* 757-765.

Holmboe, E.S., Sherbino, J, Long, D.M., Swing, S.R. & Frank, J.S. (2010). The role of assessment in competency-based medical education. *Medical Teacher, 32,* 676-682.

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *BMJ: British Medical Journal*, *311*(7001), 376.

Leighton, J. P., & Gierl, M. J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice, 26,* 3-16.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education, 35*, 501-517.

Steinert, Y. (2000). Faculty development in the new millennium: key challenges and future directions. *Medical Teacher, 22* (1), 44-50.

Verhoeven B. H., Verwijnen G. M., Scherpbier A. J. J. A., Schuwirth L. W. T. & van derVleuten C. P. M. Quality assurance in test construction: the approach of a multidisciplinary central test committee. Educ. Health 12, 49–60 (1999))

UAPPOL (2012). University of Alberta Policies and Procedures On-Line. May 28, 2012. https://policiesonline.ualberta.ca/PoliciesProcedures/Policies/Assessment-and-Grading-Policy.pdf