A UQ Assessment Brief on “Peer review improves the quality of MCQ examinations”

(Brief No: 26 - May 2012)

This month’s assessment brief is selected because of the increased interest in using Multiple Choice Questions for assessment as an administrative response to increasing class sizes. However, as these authors point out, “...there is considerable evidence that MCQs are poorly written and frequently assess recall rather than higher order thinking”. They also comment that “…poor quality items can have detrimental effects on learning and the reliability and validity of assessment”. Their paper describes the introduction of a peer review process and its sustained impact on the quality of MCQ items.

Bunmi S. Malau-Aduli & Craig Zimitat (2011, iFIRST): Peer review improves the quality of MCQ examinations, Assessment & Evaluation in Higher Education

To link to this article: http://dx.doi.org/10.1080/02602938.2011.586991

Abstract

The aim of this study was to assess the effect of the introduction of peer review processes on the quality of multiple-choice examinations in the first three years of an Australian medical course. The impact of the peer review process and overall quality assurance (QA) processes were evaluated by comparing the examination data generated in earlier years (2008) with those held under the new QA regime (2009 and 2010) from the same blueprint. Statistical analysis and comparisons of overall examination performance were made by year. Regarding multiple-choice questions (MCQs), item analysis was used to compare the proportion of difficult and discriminating items and functional distracters on summative examinations in 2008 (pre-implementation of peer review) and 2009 and 2010 (post-implementation). The impact of peer review processes resulted in a decrease in the number of items with negative discrimination; increases in reliability, appropriate item difficulty, and numbers of items with significant discrimination. There was an associated improvement in the effectiveness of distractors for the MCQ items. The trend of overall improvement in the quality of MCQ items continued in 2009 and 2010. The introduction of QA processes, specifically peer review of MCQ items has resulted in a sustained improvement in the quality of MCQ items within our examinations.

Keywords

assessment; quality assurance; multiple-choice questions; item analysis; test construction

Introduction

The introduction asserts the centrality of assessment to any evaluation of the quality of an educational enterprise and argues that the evaluation of assessment needs to occur at multiple levels, each with its own distinctive set of influences on assessment quality and student learning.

Institutional level: policies, guidelines and procedures
School and course level: blue printing, balancing formative and summative assessment strategies, choice of instruments, staff training and feedback processes

Individual level: personal choices related to the selection of subject matter, engagement in training and review an overall commitment

The paper then discusses the importance of reliability and especially validity as determinants of MCQ quality and provides some technical information about item analysis processes. The section finishes with a brief review of the literature on which the authors based their decision to trial peer review processes to enhance the quality of multiple choice examinations. Statistical analysis was employed to evaluate the outcome of the trial.

Organisational context of this study

The trial was undertaken in the University of Tasmania's medical program where a Medical Education Unit (MEU) is responsible for assessment management including undertaking a range of processes for assuring assessment quality and the maintenance of an assessment item data base. The MEU ran workshops to implement a peer review of assessment items before they were administered to students. The workshops encouraged item writers to develop MCQ items that assessed high cognitive level processes rather than recall of factual knowledge.

Methodology

The study was based on 866 items from the MEU data base and compared MCQ items developed in 2008 prior to the introduction of the peer review process with items developed in 2009 and 2010 which had been blueprinted and reviewed by colleagues and assessment committees before use. All items were five-option, single best answer questions with no penalties for incorrect answers.

Statistical analysis (which is fully detailed) indicated significant differences between the pre-and post peer review sets of MCQ items in relation to features such as item difficulty (measured by the proportion of examinees with correct responses), discrimination index (measured by the proportion of high and low achieving students with correct answers), point-biserial correlation (measured by the correlation between the scores that students receive on a given item and a total score for the examination) and functionality or plausibility of distractors (measured by the frequency with which they are selected by students).

Discussion

The study generated a range of evidence that a peer review process can improve the quality of MCQ items.

- the 2008 items have consistently higher proportions of easy, difficult, recall, non-discriminating and non-functional items than the 2009, 2010 items
- 21% of distractors in the 2008 items were not selected by any students, but that this figure had reduced to 9% in 2009 and 10% in 2010.

In addition, the low number of items in which all four distractors (incorrect answers) were found to be plausible, indicated the difficulty in creating five option sets of MCQs which suggests that achieving MCQs with three plausible distractors is a more realistic and practical approach.

In addition, outcomes of the study indicated that peer review training workshops were more productive when they focused on (1) what the question tested and its structure, (2) question relevance or importance and (3) the extent to which it addressed core knowledge, than when they focused on the minutiae of wording and formatting. Other productive steps included the
establishment of item-writing guidelines, the use of inter-disciplinary committees to review exam questions prior to their incorporation in examinations and setting and communicating clear processes such as committee review dates in a timely manner.

In summary, the authors conclude that the review process “... increased content validity of test items, reduced technical flaws and improved overall item quality because ambiguity in question interpretation was detected, multiple answers or lack of a one best answer were picked up, items tested higher cognitive learning abilities and suggestions were preferred to strengthen the test items before they were administered to the students.”

This series of briefs on assessment topics has been prepared by the UQ Teaching and Educational Development Institute (TEDI) for UQ teaching academics. “UQ ASSESSMENT BRIEFS” of journal articles, book chapters, reviews, websites, reports etc are distributed to Faculty and School Teaching and Learning Chairs in a form designed to encourage wider distribution.

Please send feedback, requests or suggestions for future “UQ ASSESSMENT BRIEFS” to:

Dr Clair Hughes (clair.hughes@uq.edu.au)
Senior Lecturer in Higher Education
Teaching and Educational Development Institute (TEDI)
The University of Queensland, St Lucia Qld 4072
Telephone +61 7 336 52456
Facsimile +61 7 3365 1966