INTRODUCTION

Multiple-choice questions (MCQs) are frequently used in medical exams because of positive psychometric characteristics, a long history of research evidence, versatility in testing more cognitive knowledge as well as ease of scoring. At the same time constructing a good MCQ is itself a laborious and lengthy exercise requiring an hour to write a good question. There are general guidelines on how to construct an MCQ (Table 1) however all MCQ formats are prone to construction errors and studies have identified flaws in MCQs.

MCQs are frequently used in online formative assessments and as part of summative assessment across the Faculty of Medicine, Dentistry and Health Sciences (FMDHS) at the University of Western Australia. These MCQs and other test items are written by the faculty members involved in teaching and are occasionally reviewed within schools responsible for that content. The Education Centre at FMDHS provides support as and when required in terms of conducting workshops and reviewing assessment results. A review of MCQs developed for the assessment of medical students was undertaken to determine the quality of multiple-choice questions and their adherence to unanimously accepted guidelines for writing MCQs. One area of the medical course was reviewed for convenience, however, the information or method of review can be applied to other areas.
METHODS

Two hundred and ninety MCQs (one best type) were reviewed to see if they follow the generally accepted guidelines suggested for their construction in the literature. The questions assessed were created for the online assessment of para-clinical sciences students in the medical course and include disciplines of Pharmacology, Pathology, Immunology, Microbiology and Infectious diseases. The review was undertaken by the author who has expertise in educational assessment and is also a qualified medical specialist. The flawed items were further discussed with a team at the Education Centre to validate the errors.

The items were classified as standard or flawed. A standard item was one that followed the guidelines for item writing while a flawed item was one that violated these guidelines. For this study, we used the guidelines validated and supported by published research studies. (Table 1)

Table 1: General guidelines for constructing a multiple choice question

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Construct each item to assess a single written objective</td>
</tr>
<tr>
<td>2.</td>
<td>State the stem in simple clear language</td>
</tr>
<tr>
<td>3.</td>
<td>Use as much as the information in stem rather than options but avoid irrelevant information</td>
</tr>
<tr>
<td>4.</td>
<td>State stem in a positive form, if negatives are to be used make them bold and/or in upper case</td>
</tr>
<tr>
<td>5.</td>
<td>Do not use all of the above as one of the options</td>
</tr>
<tr>
<td>6.</td>
<td>Use none of the above option with caution</td>
</tr>
<tr>
<td>7.</td>
<td>Do not use double negatives i.e. negative stem as well as option</td>
</tr>
<tr>
<td>8.</td>
<td>Avoid verbal clues that may lead students to select the correct choice</td>
</tr>
<tr>
<td>9.</td>
<td>Follow the normal rules for grammar and punctuation</td>
</tr>
<tr>
<td>10.</td>
<td>Place item choices alphabetically or in logical sequence</td>
</tr>
<tr>
<td>11.</td>
<td>Keep length of options equal</td>
</tr>
<tr>
<td>12.</td>
<td>Avoid use of the specific determiners such as always, never etc.</td>
</tr>
</tbody>
</table>

Of the total items reviewed, one hundred and fifty-one test items had been already administered in the examination, so we were also able to review the performance of distracters within each test item in addition to the uniform guidelines

RESULTS

PRE-EXAMINATION REVIEW

One hundred and ten questions (38%) did not follow the general guidelines for developing test items and required modification. Forty-four (12%) test items had more than one error in their construction. The errors commonly observed in writing test items are listed in Table 2 in the order of their incidence from highest to lowest.

Table 2. Frequency of common errors found in the construction of multiple choice questions (n=290 questions)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Stem structure requires rephrasing</td>
</tr>
<tr>
<td>2.</td>
<td>Unfocused stem</td>
</tr>
<tr>
<td>3.</td>
<td>Extraneous material in the stem or the options</td>
</tr>
<tr>
<td>4.</td>
<td>Negatively phrased questions, no emphasis on negatives and using double negatives</td>
</tr>
<tr>
<td>5.</td>
<td>Using all of the above as one of the options</td>
</tr>
<tr>
<td>6.</td>
<td>Only four plausible options, incorrect option standing out</td>
</tr>
<tr>
<td>7.</td>
<td>Mutually exclusive options</td>
</tr>
<tr>
<td>8.</td>
<td>Length of options not equal</td>
</tr>
</tbody>
</table>

POST EXAMINATION REVIEW

All the item (n=151) that had already been used in earlier examinations with students were also reviewed for the performance of distracters. The common flaws identified on distracter analysis were:

i. Implausible Distracters

Each of the reviewed item had five options, thus there were 755 options in all. One hundred and ten options were not selected by any of the students and 65% of these 110 options were the fifth option which even a student with minimal knowledge of the subject could have easily eliminated as a choice.

ii. More than one correct choice

Three items had more than one correct choice. In one question it was obvious as the same option was repeated twice, while in the other two items it became obvious by looking at the options which were identical to each other and that either of the two could be selected as correct choice.

iii. Distribution of correct choice

A placement bias was observed in the distribution of correct choices among five options offered (Fig.1). The correct choices are more commonly placed between options B – D. The observed frequency of correct choice in each cell was compared with the expected frequency (20%) using chi
square calculation. This yielded a value of 10.026(d.f=4) which was significant at 0.04 level.

**Figure 1. Distribution of correct answers among the five options for 290 multiple choice questions**

**DISCUSSION**

A number of flaws have been identified in this review of test items spread equally between item stems and response options.

The errors identified in the stem construction can easily be corrected; however, when we look at the flaws in options, the scenario is different. Writing five options for the MCQs has been observed as a difficult task for the Faculty in this review. This compromises the effectiveness of the distractors as it is suggested that the response rate for each distractor should be around 20 – 30 %; a distractor with low response rate does not differentiate between students who have learnt the material from those who have not learnt. It is also observed from the post-examination analysis that students most of the time are choosing from among four options only because the fifth option is immediately eliminated. Literature suggests that having three good options is better than four options of which three are good and one option is so weak that it is obviously wrong.

Correct response on an MC test should be spread among all options positions. When guessing, students select the middle options more often than extreme ones. The academics tend to hide the correct response in the middle of the options and not first or last option.

Another error that was frequently observed and is relevant to be mentioned here was the actual format of the test item itself. Some of the questions do not seem appropriate for the format of MCQ (one best type). For example, some items are testing different outcomes within one item while in others there is a repetition of the same stem in a group of test items such as ‘the mechanism of (drug name) is.....’ In such cases it is better to use a different form of objective test item like Multiple true/false type or Extended matching questions.

In late, item-writing templates called “item shells” were introduced. These facilitate the process of writing test items for novice writers. An item shell is a skeletal item containing a syntactic structure useful for writing sets of similar items. Writing high-quality test questions is a difficult task. A flawed MCQ interferes with accurate and reliable interpretation of test scores and also affects students’ passing rate by adding construct-irrelevant variance (CIV) to the assessment. Construct-irrelevant variance means that the test has variables that are irrelevant to the measured construct thus affecting itemvalidity. This may help a student respond correctly because of extraneous clues or may make it more difficult for the student to answer correctly.

It is therefore important that Faculty is trained in item writing to improve the quality of test items. Once written, these should be critically reviewed. While pre-examination review focuses on the construction, relevance and validity of the test items, the post-examination looks at how the test items have performed in the examination. Post examination assesses for items’ level of difficulty (proportion of students who respond to the item correctly), discrimination (i.e. the ability to distinguish between students who are more versus less knowledgeable) and correlation of a student’s correct response on the item with the total score.

An effective item review requires a team-based approach, which includes faculty member(s) involved in question design and selection as well as a member with expertise in test construction. The reviewed items may then be stored in item banks. An item bank is an efficient system for storage, maintenance and retrieval of test items with complete statistical information about each item and the corresponding learning outcome. With the advent of information technology, several software programs have been marketed to maintain item banks, and some institutions have developed their own. Few institutions have also collaborated jointly to share the item banks which would provide broader access to a larger pool of standardized items. It is especially feasible in health-related fields where the academics have responsibility of providing patient care along with the academic role.

**CONCLUSION**

This study was a review of multiple-choice questions at a smaller scale and looking at only one aspect of assessment using convenience sampling. The analysis found that there are many flaws in the construc-
tion of test items in our faculty, which can make the questions ambiguous or ineffective for objective assessment. This interferes with making a valid assessment of students’ learning. Therefore, in a high stakes examination, it is important to review and edit test items before and after administration.

This review has assisted the Faculty in recognising the importance of item review. It has now become common practice within the Faculty for questions of all types, not just MCQs, to be reviewed by both the discipline faculty and faculty of the Education Centre. Test Review Groups have also been established within other disciplines for interdisciplinary reviews. This not only allows the questions to be revised immediately but also nurtures a climate of peer support and a chance to look at the learning outcomes being assessed across disciplines. Regular workshops on item writing have also been introduced for faculty. These are typically scheduled during non-teaching weeks and take place at least once every semester. In addition, a lead academic in assessment provides one on one support to unit coordinators and is available when needed, which can help busy academics. These interventions have had a positive effect on the quality of test items used with our students, as evidenced by from subsequent reviews within the medical and other courses within the Faculty.

ACKNOWLEDGMENT

The author acknowledges and wishes to thank all the unit coordinators who very kindly allow their test items to be reviewed.

REFERENCES