January 2014

Evaluation of multiple choice and short essay question items in basic medical sciences

Mukhtiar Baig  
*King Abdulaziz University*

Syeda Kauser Ali  
*Aga Khan University, syeda.ali@aku.edu*

Sobia Ali  
*Liaquat National Medical College*

Nighat Huda  
*Bahria University Medical & Dental College*

Follow this and additional works at: [http://ecommons.aku.edu/pakistan_fhs_mc_cpe](http://ecommons.aku.edu/pakistan_fhs_mc_cpe)  
Part of the [Medical Education Commons](http://ecommons.aku.edu/pakistan_fhs_mc_cpe), and the [Medical Sciences Commons](http://ecommons.aku.edu/pakistan_fhs_mc_cpe)

Recommended Citation  
*Pakistan Journal of Medical Sciences, 30*(1), 3-6.  
Available at: [http://ecommons.aku.edu/pakistan_fhs_mc_cpe/10](http://ecommons.aku.edu/pakistan_fhs_mc_cpe/10)
Evaluation of Multiple Choice and Short Essay Question items in Basic Medical Sciences

Mukhtiar Baig¹, Syeda Kauser Ali², Sobia Ali³, Nighat Huda⁴

ABSTRACT

Objectives: To evaluate Multiple Choice and Short Essay Question items in Basic Medical Sciences by determining item writing flaws (IWFs) of MCQs along with cognitive level of each item in both methods.

Methods: This analytical study evaluated the quality of the assessment tools used for the first batch in a newly established medical college in Karachi, Pakistan. First and sixth module assessment tools in Biochemistry during 2009-2010 were analyzed. Cognitive level of MCQs and SEQs, were noted and MCQ item writing flaws were also evaluated.

Results: A total of 36 SEQs and 150 MCQs of four items were analyzed. The cognitive level of 83.33% of SEQs was at recall level while remaining 16.67% were assessing interpretation of data. Seventy six percent of the MCQs were at recall level while remaining 24% were at the interpretation. Regarding IWFs, 69 IWFs were found in 150 MCQs. The commonest among them were implausible distracters (30.43%), unfocused stem (27.54%) and unnecessary information in the stem (24.64%).

Conclusion: There is a need to review the quality including the content of assessment tools. A structured faculty development program is recommended for developing improved assessment tools that align with learning outcomes and measure competency of medical students.

KEY WORDS: Assessment, MCQ, SEQ, Item analysis.

Abbreviations Used: Multiple Choice Questions (MCQs), Short Essay Questions (SEQs), Item Writing Flaws (IWFs).

doi: http://dx.doi.org/10.12669/pjms.301.4458

How to cite this:

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Assessment is an essential part of the learning process in education. Students perceive it as a dominant motivator to direct and drive their learning.¹ The method of assessment determines the approach of students towards learning. Students’ are inclined to espouse a surface approach when assessment emphasis is on recall of factual knowledge and students are more likely to adopt a deep approach² if assessment demands higher levels of cognitive abilities. The approach to learning is a dynamic characteristic and is always modified according to students’ perceptions of the learning environment.³ It has been reported that one of the most important factor influencing students’ choice of learning approach is the way how assessment is being conducted.⁴⁻⁶
Multiple methods of assessment namely MCQs, SEQs, OSPE and VIVA are commonly used to assess Basic Science knowledge in undergraduate medical education in Pakistan. Multiple choice questions (MCQs) are the most frequently used type of tests deployed on their own or in combination with other types of test tools for assessment. Moreover, MCQs are appropriate for measuring knowledge, comprehension and could be designed to measure application and analysis. MCQs are being used increasingly due to their higher reliability, validity, and ease of scoring. Essay-type assessment is a sensitive test requiring students not only to recall facts but also to use higher-order cognitive skills. Essay questions though time consuming provides a unique evaluation tool particularly suited for the undergraduate settings.

The use of multiple formats is recommended in assessment of medical students. However, assessment tools should be valid and reliable and be able to measure the different aspects of professional competencies. The present study was undertaken to evaluate MCQ and SEQ items in Basic Medical Sciences (Biochemistry) by determining item writing flaws (IWFs) of MCQs along with cognitive level of each item in both methods.

**METHODS**

This analytical study was carried out in the department of Biochemistry, in a newly established medical college in Karachi, Pakistan. The first batch was admitted in January 2009 while the undergraduate curriculum has been organized in six limited integrated modules. Multiple assessment methods including short essay questions, MCQs, OSPEs, and orals carrying equal weightage were used for module assessment. The college faculty developed all assessment questions.

The cognitive levels of the assessment tools were analyzed using the Buckwalter’s (Buckwalter et al. 1981) modification of the Bloom’s taxonomy (Bloom 1956). Level I: Include questions which attempt to check recall of information. Level II: Include questions which attempt to test understanding and interpretation of data. Level III: Include questions which attempt to test the application of knowledge for solving a particular problem.

For determining types of item writing flaws (IWFs) standard criteria given by Case and Swanson (2003), Haladyna et al., (2002) & Tarrant et al., (2008) were used and 14 commonly occurring violations of item-writing guidelines were identified. One subject expert and three medical educationists reviewed each assessment tool. Initially the reviewers individually reviewed the assessment tool for the cognitive level and IWFs according to predefined criteria and reported their results to the principal investigator. A consensus meeting was called to reach on unanimous decision about the debatable questions.

**RESULTS**

A total 150 of MCQs that were administered in all six module examinations were reviewed. The cognitive level of 114 MCQs (76%) was at recall level while remaining 36 MCQs (24%) were of interpretation of data and there was no MCQ evaluating problem solving domain of knowledge (Table-I). A total of 36 SEQs were administered in all six module examinations. The cognitive level of 30 SEQs (83.33%) was assessing recall of knowledge while remaining 6 SEQs (16.67%) were assessing interpretation of data and there was no SEQ assessing problem solving domain of knowledge (Table-I).

A total of 150 MCQs were administered in the all six module examinations. Upon review, 69 IWFs were found and four most common IWFs were implausible distracters (30.43%), unfocused stem (27.54%), unnecessary information in the stem (24.64%), and negative stem (8.7%), respectively (Table-II).

**DISCUSSION**

The selection of an appropriate assessment method for measuring students’ performances remains a daunting task for many medical institutions in Pakistan. Attempts to change existing assessment

---

Table-I: Distribution of cognitive levels of SEQs and MCQs in all six modules. (n= Number of SEQs, and MCQs in each module)

<table>
<thead>
<tr>
<th>Module</th>
<th>Cognitive levels of SEQs (n=6)</th>
<th>Cognitive levels of MCQs(n=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1</td>
<td>C2</td>
</tr>
<tr>
<td>M1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>M2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>M3</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>M4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>M5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>M6</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

| Total  | 31 | 5  | 0  | 114| 36 | 0  |

| Percentage | 86.11% | 13.89% | 0% | 76% | 24% | 0% |

C1= Cognitive level 1, C2= Cognitive level 2, C3= Cognitive level 3
methods have been hindered by financial constraints, lack of expertise in psychometric analysis of the examination and institutional policies.

The present research found that 76% MCQs were testing the recall of isolated facts while remaining 24% MCQs were testing the skill of interpretation of data and there was not a single MCQ assessing the higher cognitive domains of application and analysis. It may be due to the fact that MCQs at recall level are easier to construct and need less time and knowledge as compared to problem-solving MCQs which need expertise and training.8,9

Khan and Aljarallah (2011)17 in their study found that the percentage of MCQs testing the level III (problem solving) cognitive skills of the students was 60%, level II (interpretation of data) was 6% and level I recall of information was 28%. But in that study a total of 50 MCQs representing different disciplines of medicine like gastroenterology, cardiology, neurology, rheumatology, nephrology etc were evaluated. Constructing problem solving MCQs in basic sciences is difficult in comparison to clinical sciences. In present study, multiple factors have contributed to low cognitive level questions such as newly established medical college with inadequate faculty training, diverse background of teachers, non existence of question bank, first examination of the college etc. Tarrant and Ware (2008)16 found in a nursing examination that over 90% of MCQs were written at low cognitive levels, and that MCQs written at a lower cognitive level were significantly more likely to contain item-writing flaws. Jozefowicz et al. (2002)18 evaluated the quality of in-house developed examinations at three US medical schools and found that the overall quality of the questions used was low. Several studies have confirmed that MCQs not only test the knowledge of the examinees but can also be used for measuring higher cognitive skills.19,20

One of the most common problems affecting MCQs quality is the presence of item writing flaws. Item-writing flaws (IWFs) are violations of these accepted item-writing guidelines which can affect student performance on MCQs, making the item either easier or more difficult.21 The present study found 69 IWFs (46%) in total 150 MCQs in Biochemistry module exams, and four most common IWF were implausible distracters (30.43%), unfocused stem (27.54%), unnecessary information in the stem (24.64%), and negative stem (8.7%). These results are in agreement with several studies.16,20,22,23

Another study documented that 12% of their exam MCQs had item writing flaws.17 But in that study all MCQs belonged to clinical sciences subjects, and 60% of the MCQs assessed students’ problem solving skills. In our study problem solving MCQs were zero% and 76% of the MCQs assessed recall of isolated facts. Our higher percentages IWFs can be explained in the view of Tarrant and Ware (2008) study who concluded that MCQs written at lower cognitive levels, are more likely to contain item-writing flaws.16 Downing (2005)21 assessed the quality of four examinations given to medical students in the United States of America, and found that 46% of MCQs contained IWFs and reported that as a consequence of these IWFs, 10–15% of students who were classified as failures would have been classified as pass if items with IWFs were removed. Results of present study showed presence of 46% of flawed items which could be mostly attributed to insignificant faculty development programs. Flawed items affect difficulty and discrimination index. Low difficulty and poor discrimination in an item favors low achievers while higher difficulty

<table>
<thead>
<tr>
<th>Types of IWFs</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute terms</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vague terms</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Implausible distractors</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>21</td>
<td>30.43</td>
</tr>
<tr>
<td>Extra details in correct option</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.44</td>
</tr>
<tr>
<td>Unfocused stems</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>19</td>
<td>27.54</td>
</tr>
<tr>
<td>Grammatical clues</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Logical clues</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4.55</td>
</tr>
<tr>
<td>Word repeats</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.44</td>
</tr>
<tr>
<td>&gt; 1 correct answer</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unnecessary information in stem</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>17</td>
<td>24.64</td>
</tr>
<tr>
<td>Lost sequences in data</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.44</td>
</tr>
<tr>
<td>All of the above</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>None of the above</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Negative stem</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>8.7</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>18</td>
<td>10</td>
<td>07</td>
<td>05</td>
<td>08</td>
<td>69</td>
<td></td>
</tr>
</tbody>
</table>

Results of present study showed presence of 46% of flawed items which could be mostly attributed to insignificant faculty development programs. Flawed items affect difficulty and discrimination index. Low difficulty and poor discrimination in an item favors low achievers while higher difficulty
and poor discrimination negatively affected the high scorers and moreover flawed items also fail to assess the course learning objectives.23

For reducing IWFs, and improving cognitive levels of the test items, Downing 2006, suggested the use of test blue print.24 A blue print is simply a grid or table that maps the course objectives and content to be tested, and is an essential step in generating a valid and reliable test. Test blue print helps in accurately delineating the percentage of test questions to be allocated to the different content areas, and at different cognitive levels.

Faculty should be encouraged, and trained to construct MCQs for higher order cognitive levels. Tarrant et al., (2006)25 pointed out that by removing IWFs from MCQs does not necessarily change the cognitive domain of a question, but writing questions at higher cognitive levels inherently remove numerous IWFs.

The present study suggest that there is need to improve the quality of our assessment tools because if the assessment tools measure low cognitive level, it will not only decrease the validity of the exam but also compel the students to adopt surface learning approaches which is not suitable for lifelong learning.

Limitations of the study: The study analyzed results of only two modules, and students’ scores in only one subject. Moreover, difficulty and discrimination indices were not available.

CONCLUSION

The medical college should evolve policy guidelines on preparing questions of higher cognitive level for all departments and student assessment should align with the learning outcomes. There is need to review the quality of item including the content of assessment tools.

Recommendations: Psychometric analysis should be done for all types of assessment while test blue print should be developed to ensure validity. A structured faculty development program is recommended for developing improved assessment tools. Similar studies are recommended in other basic science subjects.

REFERENCES


Authors Contribution:
MB: Designed the study, did all analysis and prepared the manuscript.
KA, SA, NH: Helped in evaluating the quality of the assessment tools, drafting and revising the manuscript.