Psychometric Analysis of MCQs Used in Assessing the Students at Entrance to a Medical College

Syed Shoaib Hussain Shah,[1] Tahir Ahmad Munir,[2] Muhammad Sabir,[3] Salman Ahmad Tipu[4]

Abstract

Objective: The study was performed to investigate the reliability and validity of MCQs paper administerred to the applicants of the first MBBS batch at Rawal Institute of Health Sciences Islamabad.

Study Design: Analytical study.

Material and Methods: 235 students were administered a questionnaire consisting of 100 MCQs of single best type with 5 options at the entrance to Rawal Institute of Health Sciences, Islamabad. The MCQs were generated by the subject specialists, vetted at departmental and central levels and were analyzed in terms of its reliability, validity and difficulty with discriminating indices. For reliability in terms of internal consistency Cronbach's alpha was used and was 0.72. Difficulty and discriminating indices for the MCQs were collected from the computer generated marking sheets.

Result: The MCQs showed satisfactory levels of reliability and validity and majority of the MCQs were within the acceptable level of difficulty index. A well-structured and strict central vetting process ensures an acceptable standard MCQs.

Conclusion: MCQ testing, the most efficient form of written assessment is both reliable and valid. Such testing of cognitive knowledge predicts and correlates well with overall competence and performance. MCQ 'fairness' is an increasingly important strategic concept to improve their validity.

Keywords: Psychometric characteristics, MCQs, Medical students.

Shah S.S.H.[1]
Department of Forensic Medicine, Rawal Institute of Health Sciences, Islamabad

Munir T.A.[2]
Departments of Physiology / Medicine, Rawal Institute of Health Sciences, Islamabad

Sabir M.[3]
Department of Examination, Al-Nafees Medical College, Islamabad

Tipu S.A.[4]
Department of Medical Education, Rawal Institute of Health Sciences, Islamabad

Introduction

Psychometric analysis can be applied to improve or validate almost any instrument that measures mental performance and identify personality and potentials abilities. Over the last decade, larger student numbers, reduced resources and increasing use of new technologies had led to increase use of multiple-choice questions (MCQs) as a method of assessment for admission in higher education courses in almost every part of the world.[1] Entrance examination / assessment plays an increasing role in satisfying quality issues of registration and examining bodies as well as reassuring the public.[2]

A longstanding criticism of the validity of MCQs is that, testing cognitive (or factual) knowledge does

not guarantee competence as professional competence integrates knowledge, skills, attitudes and communication skills.[2] However, decades of research into reasoning and thinking have unequivocally shown that knowledge of a domain is the single best determinant of expertise. MCQs are, therefore, a valid method of competence testing, as cognitive knowledge is best assessed using written test forms.[2,3] While MCQs are expressly designed to assess knowledge, well – constructed MCQs can also access taxonomically higher – order cognitive processing such as interpretation, synthesis and application of knowledge rather than testing recall of isolated facts.[4] Over – reliance on written forms of assessment can lead to unforeseen, unwanted educational consequences such as over-reliance on written learning. To make testing both fair and consequentially valid, MCQs should be used strategically to test important content, and mixed with practical testing of clinical competence. Though often maligned and true that no single format should be used exclusively for assessment, American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education recommend MCQs testing as one of the most commonly used formats both for formative and summative assessments.[5]

This study was performed to investigate the reliability and validity of MCQs paper administered to the applicants of the first MBBS batch at Rawal Institute of Health Sciences Islamabad.

Methodology

A total of 235 students appeared in the admission test of first MBBS batch at Rawal Institute of Health Sciences Islamabad. They attempted 100 MCQs each with 05 one best type responses. Marks obtained by the individual student were collected and were analyzed by SPSS version 16.0, while reliability in terms of internal consistency, Cronbach's alpha were used.

For item analysis, the response sheets were scored on optical mark reader version 8.0. Two important indices provided by the item analysis were the difficulty index / factor and the discrimination index.

Difficulty index was calculated as DI: Number of students who answered the test item correct to total Number of students who attempted the item× 100. Discrimination Index or power, is a measure of how well the item differentiates among high and low scoring students. Either way, it was expressed on a scale from

-1.00 to +1.00. Negative 01 means all low scorers who got the item right and all high scorers who got the item wrong. A positive 01 means the item worked exactly as it should. A zero means the item didn't distinguish between good and bad students.

To find out how hard or easy was the item for the students, was computed as the proportion of students who got the item correct. A low value means a hard question and a high value means an easy question on a scale from 0.00 to 1.00.

Results

The results were classified on the basis of point – Bi-serial and P-value. Easy items showed a P value > 0.69 and point – Bi-serial ranging from 0.01 – 0.20; moderate items had P-value ranging from 0.30 – 0.69 and point – Bi-serial 0.21 – 0.50; while hard items had a P-value < 0.30 and point Bi-serial > 0.50. Few items also showed negative point – Bi-serial (Fig. 1, 2). The result showed 48% easy items, 22% moderate and 21% hard items, while 09% of the items were of negative point –Biserial (Fig. 3). The internal consistency of the whole sample was 0.72.
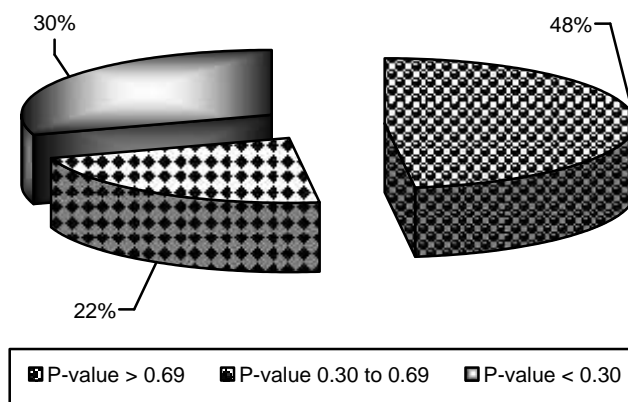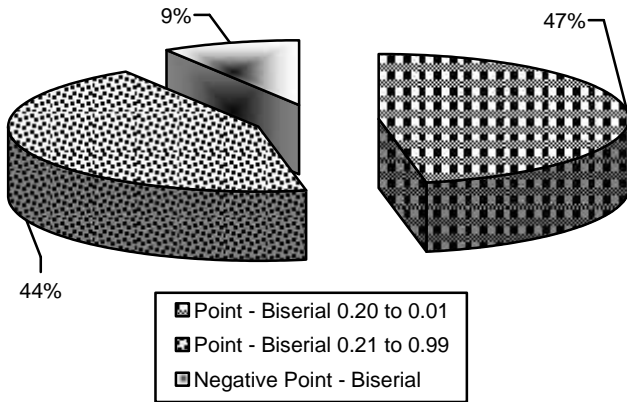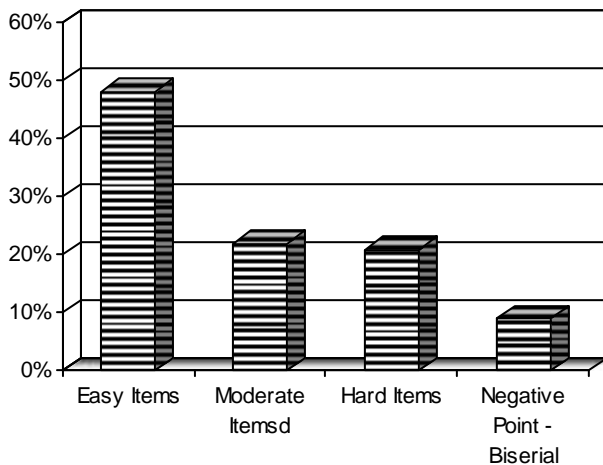

Fig. 1: Percentage Distribution of P-Value.

Discussion

The MCQs used in the entry test exam to select the students for admission at Rawal Institute of Health Sciences Islamabad were reliable, easy to mark and to sample a large part of the curriculum. There has been, and will continue to be, much debate about the use of multiple choice questions (MCQs) for the assessment of medical students. A longstanding criticism of the

Fig. 2: Percentage Distribution of Point – Bi-serial.

Fig. 3: Percentage Distribution of MCQs.

validity of MCQs is that testing cognitive (or factual) knowledge does not guarantee competence as professional competence integrates knowledge, skills, attitudes and communication skills.[6] To say that MCQs mainly test factual recall is not a fault of MCQs per se, but a reflection of the way in which such questions are constructed, so it is possible to devise good MCQs with considerable effort which can test higher cognitive skills, attitudes and communication skills. However, decades of research into reasoning and thinking have unequivocally shown that knowledge of a domain is the single best determinant of expertise.[6] MCQs are, therefore, a valid method of competence testing, as cognitive knowledge is best assessed using written test forms.[7]

Another criticism of using MCQs to assess the medical students is the choice of a small number of alternatives that may provide clues to the answer which the candidates would not have generated if left to their own devices (Cueing); as patients do not usually present with a list of five alternative treatments or diagnoses and, in real life, clinicians are required to generate their own options, cueing reduces the validity of MCQs for assessing the application of knowledge in areas such as diagnosis and treatment.[8] Evidences exist that these skills could be assessed by using short answer questions that provide one approach if cueing is a concern[9,10] and considered to fall in between MCQs and essays.[1]

Our results showed satisfactory level of reliability and validity. The reliability coefficient for internal consistency using Cronbach's alpha was 0.72 for the whole sample (range 0.60 and 0.85) and based on number of test items and use of the results. The Face and content validity measure how well the test items represent the domain of learning objectives.[10,11] There is no statistics to establish the content validity except that how the MCQs were designed and evaluated by the expert faculty so the consensus development techniques justify both face and content validity in our case.

The difficulty index in our case was < 0.30 to > 0.69. The difficulty index is not solely determined by the content of the item but also reflects the ability of the examinee[12] and the instructions they have had.[13] For a well – prepared group of examinees item difficulty indices may range from 70 to 100%. In the entry test exam 93.3% pass rate indicates a well – prepared group of students and this may be the reason of high difficulty indices for some of the test items. A rigid content specification should be maintained in generating the items and for that purpose, items with high difficulty indices may need to be accepted.[12]

Discriminating index refers to the degree to which the test item discriminates between students with high and low achievement. When the difficulty index moves towards high or low from 50%, the discriminating index becomes low.[13] Si-Mui et al[14] showed that MCQ items with good discriminating potential tend to be moderately difficult items, and the moderate to very difficult items were more likely to have negative discrimination. Preferable discriminating indices are 0.20 and above,[15] but in criterion-referenced measurement, many good items may have discrimination indices of zero.[12] True – false format MCQs provide cues, resulting in a less discriminatory index.[16] Again low discriminating index is more likely if the test measures a variety of types of learning outcomes. For validity, a

well – constructed test accepts items with low discriminating indices.[13] The so – called 'assessment by ambush' is one aspect of unfair examination, where, for high discrimination, potentially important areas are not tested.[17]

Conclusion

MCQ testing is the most efficient form of written assessment, being both reliable and valid by broad over age of content. Such testing of cognitive knowledge predicts and correlates well with overall competence and performance. However, MCQs are not without validity problems. MCQ 'fairness' is an increasingly important strategic concept to improve the validity of their use

References

1. David N. E-assessment by design: using multiple – choice tests to good effect. *Journal of Further and Higher Education* 2007; 31: 53–64.
2. P.Mc – Coubrie. Improving the fairness of multiple-choice questions: a literature review. *Med Teach* 2004; 26: 709–712.
3. Downing SM. Assessment of knowledge with written test formats, in: G. Norman, C. Van D. Newble (eds) *International Handbook of Research in Medical Education.* 2002; 2: 647–72.
4. Case SM, Swanson BD. Constructing Written Test Questions for the Basic and Clinical Sciences, 3rd ed. Philadelphia, *National Board of Medical Examiners,* 2001.
5. Dianne S, Linda C, Fritz D, Brian G, Laura H, Jo – Ida H, et al. *The Standards for Educational and psychological Testing*: 2009. AERA Publications Sales, 1430 K Street, NW, Suite 1200, Washington, DC 20005.
6. Glaser R. Education and thinking: the role of knowledge, *American Psychology*. 1984; 39: 193–202.
7. Downing SM. Assessment of knowledge with written test formats, in: G. Norman, C. Van – Der, D. Newble. 3rd ed. *International Handbook of Research in Medical Education,* Dordrecht, Kluwer, 2002; Vol. 2: pp. 647–72.
8. SL. Fowell, JG. Bligh. Recent developments in assessing medical students. *Postgrad Med J*. 1998; 74: 18–24.
9. Maguire T, Skakun EN, Triska OH. Student thought processes evoked by multiple choice and constructed response item. *The Seventh Ottawa Conference on medical education and assessment.* Maastricht, 1996: 253.
10. Buckley – S, FTC Harris. An assessment of a final qualifying examination online: *BMC med Edu* 2009.
11. A. Barman, R.Jaafar, FA. Rahim, AR. Noor. Psychometric Characteristics of MCQs, used in phase II. *The Open Medical Education Journal* 2010; 3: 1-4.
12. Ebel RL, Frisbie DA. Essentials of educational measurement. 5th ed. New Jersey: Printice – Hall, Inc. 1991.
13. Linn RL, Gronlund NE. Measurement and assessment in teaching. 8th ed. New Jersey: Printice-Hall, Inc. 2000.
14. Si – Mui S, Isaiah R. Relationship between item difficulty and discrimination indices in true / false – type multiple choice questions of a para-clinical multidisciplinary paper. *Ann Acad Med* 2006; 35: 67-71.
15. Dixon RA. Evaluating and improving multiple-choice papers: True – false questions in public health medicine. *Med Educ* 1994; 28: 400-8.
16. Veloski J, Rabinowttz H, Robeson M, Young P. Patients don't present with five choices: an alternative to multiple – choice tests in assessing physicians' competence. *Acad Med* 1999; 74: 539-46.
17. Brown B. Trends in assessment. In: Harden R, Hart I, Mulholland H, Eds. Approaches to the assessment of clinical competence. UK: *Dundee Centre for Medical Education* 1992; Vol. 1.