ITEM ANALYSIS TO IDENTIFY QUALITY OF MULTIPLE CHOICE QUESTIONS AT KHYBER GIRLS MEDICAL COLLEGE

Afshan Ahsan, Naheed Mahsood, Naveed Afzal Khan

Introduction: Multiple choice questions (MCQs) are frequently used as assessment tool in medical education. Item analysis is the method used to analyze the quality of MCQs. Difficulty index, discrimination index and distractor efficiency are used to evaluate the standard of MCQs. The objective of this study was to identify the quality of MCQs currently being used in the internal assessments at Khyber Girls Medical College (KGMC).

Material and Methods: The study was conducted at the department of Medical education of Khyber Girls' Medical College Peshawar after taking approval from Dean. The Pre-Professional assessment MCQs of four departments of Third and Fourth Year MBBS were used for analysis. A total of 165 MCQS were selected for the item analysis. Each type of an MCQ contained a stem and four or five options including the key. There was no negative marking. Each item was analyzed for difficulty index, discrimination index, and distractor efficiency. The MCQs items were marked and analyzed through OMR and MS Excel. The relationship of distractor efficiency and discrimination index was determined by Pearson correlation analysis using SPSS 20.0

Result: More than 50% of the MCQs were in acceptable category. While in subject C 40% MCQs were in too difficult category. Fifty-three percent MCQs of subject D and 45% MCQs of subject C were in poor discrimination power category. There were 39% "poor DI items" while 35% MCQs had 2 or more inefficient distractor, rest of the distractors were plausible. There was strong correlation between the number of nonfunctioning distractors and difficulty index ($r^2 = 0.617$, $p<0.001$). The reliability of subject A, B, C and D was 0.74, 0.81, 0.49 and 0.50 respectively.

Conclusion: Item analysis is a simple and valuable post examination method providing information on quality of MCQs by calculating difficulty index, discrimination index and distractor efficiency. The result of this study showed high proportion of difficult items than the normal recommended. There was need to work on "poor DI items" and inefficient distractors. The result of this study would help to initiate a change in the construction and selection of MCQs and would focus on quality of assessment strategy as part of curriculum development.

Keywords: Difficulty index, discrimination index, distractor efficiency, item analysis. Multiple choice questions.

INTRODUCTION

Assessment is an integral component of any curriculum. Numerous tools are available for the purpose of assessment. The most important characteristics of any assessment tool are its content validity and reliability. Validly can be described as the ability of an assessment tool to measure what it is supposed to measure.[1] It depends on the extent an assessment tool samples the range of learning domains that are covered by the students during a specified period. Reliability is related to the consistency of the assessment. A good reliable assessment consistently achieves the same results with multiple exposures to the same (or similar) cohort of students.[1] the reliability of a test depends upon grading consistency and discrimination between students of differing performance levels. [2]

Multiple choice questions are frequently used as an assessment tool in medical education. The strength of the MCQs lie in their objectivity. A high number of MCQs can be incorporated in a single test and thus a wide range of learning outcomes can be assessed. They can be used to assess the "knows" and "knows how" level of the Millers' pyramid of assessment. [3] The level they assess is determined by the construction and quality of the MCQ. With the advent of optical mark recognition (OMR) machine the marking of MCQs has become very simple and time efficient.[4] Item analysis is a method which assesses the quality of MCQs by assessing student's responses to each MCQ.[5] Item analysis includes Difficulty index, Discrimination index and Distractors efficiency. Difficulty index describe the difficulty level of MCQs. The range of Difficulty index is from 0 to 100%. High percentage indicates the easy item and low percentage indicates a difficult item.[6, 7]

Discrimination index distinguish between high and low score students.[6, 8] Poor discrimination or negative discrimination question item should be revised as it indicates that the low scoring students opted the key more than the high scorers. The discrimination index

[1]Academic Curriculum Coordinator, Khyber Girls Medical College, Peshawar, Pakistan.
[2]Assistant Professor, Medical Education Department, Khyber Girls Medical College, Peshawar, Pakistan
[3]Lecturer, Khyber Girls Medical College, Peshawar, Pakistan

Address for correspondence:
Afshan Ahsan
Academic Curriculum Coordinator, Khyber Girls Medical College, Peshawar, Pakistan.
Email: mountainfly87@yahoo.com
Cell No: 0333-9928187

range is between -1 and +1. The closer the index is to +1, the more effectively the item distinguishes between the high and low scorers. Distractor Analysis is important in assessing the efficiency of the distractors in a test item. For distractors to be acceptable it should be opted by at least 5% candidate.[6, 9] We conducted this study to analyze the quality of MCQs currently being used in the internal assessment exams at Khyber Girls Medical College (KGMC). The purpose of the study was to provide a base line evidence of the current assessment practices and identify areas of improvement to create a viable question bank for future use.

OBJECTIVES

To determine the item difficulty index and discrimination index of individual MCQ item.To identify the frequency of Nonfunctioning distractors. To determine the correlation of Non Functioning Distractors of an item and its Difficulty index

METHODS

The study was conducted in the department of Medical education of Khyber Girls' Medical College Peshawar after taking approval from Dean. The duration of the study was two months from Sep 2017 to Oct 2017. The Pre-Professional assessment MCQs of four departments of Third and Fourth Year MBBS were used for analysis. Subject A and B MCQs paper contained 50 MCQs while subject C and D paper consist of 20 and 45 MCQs respectively. So a total of 165 MCQs were included in this study. Each type of a MCQ contains a stem and four or five options including one correct. A correct answer was given 01 mark and there was no negative marking. The MCQs items were marked and analyzed through OMR and MS Excel. The characteristic of Pre-Prof examination were given in table 01. Table 1 shows the numbers of students (Appeared, Pass, failed) and their scores (mean, minimum and maximum) for each of the 4 subjects.

RESULTS

The Reliability of the test was measured by Cronbach Alpha. The reliability of subject A, B, C and D was 0.74 , 0.81 , 0.49 and 0.50 respectively. "An item was considered difficult when the difficulty index was less than 30% and considered easy when the index was more than 70% and the value between 30-70% was acceptable. Item with negative discrimination index (D) were considered defective item; D: <0-0.019 poor discrimination item; D: 0.2-0.29; Acceptable discrimination item; D: 0.3-0.39 were good discrimination item; D : >0.4 were considered excellent discrimination item."[6] Figure 1 shows the difficulty index of Subject A, B, C and D pre-prof exams. More than 50% of the MCQs of subject A, B and D were in acceptable category. In Subject C 40% MCQs were in too difficult category. As shown in table 2, 48% MCQs of Subject A have good or excellent discrimination power. 60% MCQs of Subject B were in good or excellent discrimination power category. 45%pf Subject C and 53% MCQs of Subject D were in poor discrimination power category. Figure 2 shows the Distribution of Subject A MCQs based on the number of Non Functioning Distractors. In Subject A 70% MCQs were having 0 or 01 NFDs and 06% MCQs were having 03 or 04 NFDs. Figure 3 show the Distribution of Subject D MCQs based on the number of Non Functioning Distractors. The analysis shows that 56 % MCQs were having 0 or 01 NFDs while 02% MCQs were having 0 NFDs.

Figure 4 show the Distribution of Subject C MCQs based on the number of Non Functioning Distractors. In Subject C 40% MCQs were having 0NFDs and 10% MCQs were having 03 or 04 NFDs. Figure 5 shows the Distribution of Subject D MCQs based on the number of Non Functioning Distractors. The analysis in the Pie chart highlighted that 68 % MCQs were having 0 or 01 NFDs. From the Correlations table, it can be seen that there is a significant strong correlation between the No of nonfunctioning distractors and difficulty index(r^2=

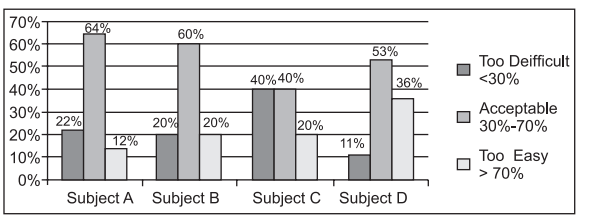

Figure 1: Difficulty Index of MCQs

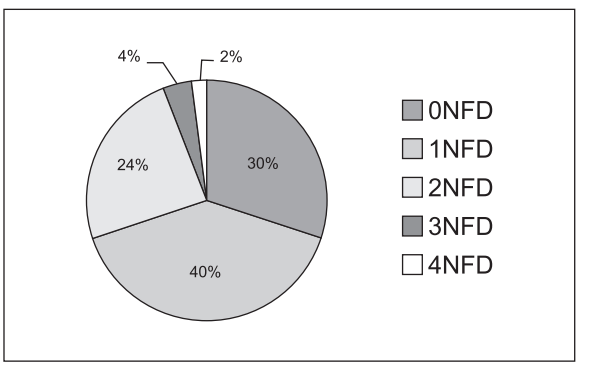

Figure 2: Distribution of Subject A MCQs based on the number of Non Functioning Distractors

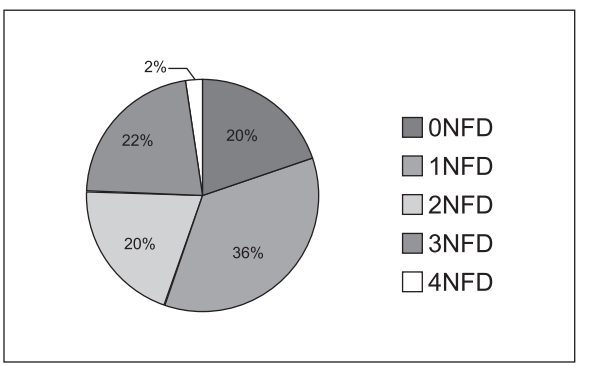

Figure 3: Distribution of Subject B MCQs based on the number of Non Functioning Distractors

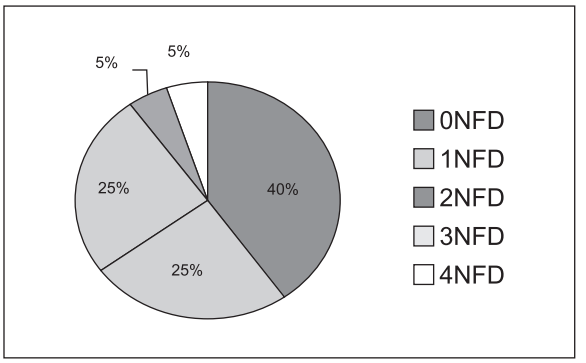

Figure 4: Distribution of Subject C MCQs based on the number of Non Functioning Distractors

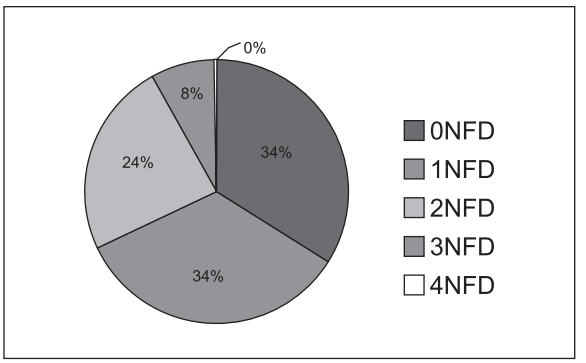

Figure 5: Distribution of Subject D MCQs based on the number of Non Functioning Distractors

0.617, p<0.001).

DISCUSSION

MCQs if properly constructed objectively assess any level of cognition according to blooms taxonomy. Analysis of MCQs after examination helps to analyze the quality of MCQS and exam paper as a whole. It also helps in identifying the content of syllabus which needs more attention and clarity and change in teaching methodology. Assessment tool should be designed according to learning objectives otherwise test will be a failure.[10] Single correct response type MCQ is an excellent tool for assessment if constructed properly and checked by item analysis.[11]Difficulty and discrimination indices are the tools that are used to check the quality of MCQs.[11]

Another tool is the distractor efficiency which analyze the quality of a distractors in MCQS. A distractor opted by less than 5% of the students inefficient distractor and should be revised and reviewed as it can affect the quality of MCQ.[12]

In our study we did item analysis of four subjects Third and Fourth Year MBBS Pre Professional MCQs paper. More than 50% of the MCQs of Subject A, B and D were in acceptable category of difficulty index. While in Subject C 40% MCQs were in too difficult category which is more than the suggested limit in literature and

Table 1: Characteristics of Examination.

Subjects	Number of MCQs	Number of Students (%)			Score (out of 100)		
		Appeared	Passed	Failed	Mean	Maximum	Minimum
Subject A	50	68	26(38)	42(62)	23	37	9
Subject B	50	69	43(62)	26(38)	25	41	8
Subject C	20	66	12(18)	54(82)	8	14	4
Subject D	45	66	56(85)	10(15)	26	41	18

Table 2: Discrimination Index of MCQs.

Subjects	Defective item (Negative)	Poor Discrimination (<0-0.19)	Acceptable Discrimination(0.2-0.29	Good Discrimination(0.3-0.39)	Excellent Discrimination(>0.4)
Subject A	**2(4%)**	13(26%)	11(22%)	5(10%)	19(38%)
Subject B	4(8%)	10(20%)	6(12%)	5(10%)	25(50%)
Subject C	2(10%)	7(35%)	5(25%)	2(10%)	4(20%)
Subject D	1(2%)	23(51%)	8(18%)	6(13%)	7(16%)

Table 3: Correlation of No of NFDs & Difficulty index.

Correlations			
		Difficultyindex	No of NFDS
Difficulty index	Pearson Correlation	1	.617**
	Sig. (2-tailed)		.000
	N	165	165
No of NFDS	Pearson Correlation		1
	Sig. (2-tailed)	.000	
	N	165	165

" Correlation is significant at the 0.01 level (2-tailed)."

resulted in highest number of failure (82%).This could be due to poor understanding of a topic or lack of clarity in wordings of the questions or even wrong key or variation in student's intelligence level. 53% MCQs of Subject D and 45% MCQs of Subject C were in poor discrimination power category. A similar type of study reported that too difficult or too easy items poorly discriminated. According to a study by Sushma S. Pande,[6] items that poorly or negatively discriminate should be reviewed for possible corrections or reconstruction. However the sample size in their study was quiet small (50) and in our study it was 165.

Very difficult and very easy items need to be revised. Item analysis serves a helpful tool in development of question bank at departmental and institutional levels which can provide MCQs with acceptable difficulty and discrimination indices.

It was also found in this study that NFDs can affect the assessment quality by decreasing the level of difficulty of the MCQs. There was strong correlation between the number of nonfunctioning distractors and difficulty index($r = 0.617$, $p < 0.001$). "Our study supported the findings of a similar study which showed that the MCQs with more NFDs are easier than the MCQs with lesser NFDs."[13] Other studies showed that the number of NFDs adversely affects the difficulty of MCQs.[11, 12, 14] Other item writing flaws like unfocused or negative stems, unnecessary information, and inefficient distractors can affect the students' performance.

Conclusion and Recommendations

Item analysis is a simple and valuable post examination method providing information on quality of MCQs by calculating difficulty index, discrimination index and distractor efficiency. It is suggested in a study that items with moderate difficulty, high value of discrimination index and with at least 3 functional distractors are of good standard to be kept for future exam.[11] The result of this study showed high proportion of difficult items than the normal recommended, items with moderate level were lesser than recommended level. There were 39% "poor DI items" while 35% MCQs had 2 or more inefficient distractor, rest of the distractors were plausible.

By item analysis, the item writer can improve the quality of an MCQ. Therefore it is recommended as an important quality assurance process before banking MCQs for assessment in future. Moreover items with negative DI must be removed or revised and NFDs should be replaced with ideal/plausible distractors. Further studies and researches are recommended for continuous check and improvement of current assessment tools.

REFERENCES

1. Carmines EG, Zeller RA. Reliability and validity assessment. Sage publications; 1979 Nov 1.
2. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian. 2005 Jan 1;12(1):19-24.
3. Cruess RL, Cruess SR, Steinert Y. Amending Miller's pyramid to include professional identity formation. Academic Medicine. 2016 Feb 1;91(2):180-5.
4. Taylor GS, inventor; Tms, Inc., assignee. Method of optical mark recognition. United States patent US 6,741,738. 2004 May 25.
5. Mehta G, Mokhasi V. Item analysis of multiple choice questions-an assessment of the assessment tool. Int J Health Sci Res. 2014;4(7):197-202.
6. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology.
7. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1, multidisciplinary summative tests. IeJSME. 2009;3(1):2-7.
8. Johnstone AH. LTSN Physical sciences practice guide: effective practice in objective assessment. Hull, LTSN. 2003.
9. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Medical Education. 2009 Jul 7;9(1):40.
10. Sudha R and Prema Janardan,. Item Analysis - The Steps and Interpretation. International Journal of Comprehensive Nursing 2014; 1(7)
11. Gyata Mehta,Varsha Mokhasi. Item Analysis of Multiple Choice Questions- An Assessment of the Assessment Tool. International Journal of Health Sciences and Research 2014; 4(7): .
12. Marie Tarrant,James Ware and Ahmed M Mohammed. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Medical Education 2009; 9(1):
13. Mufti TS, Ahsan A. Effect of non-functioning distractors on difficulty index of Physiology' MCQ bank at Rehman Medical College, Peshawar, KP, Pakistan. J Rehman Med Inst. 2015 Jul-Dec; I(2):9-13.
14. Abhijeet S. Ingale1, Purushottam A. Giri, Mohan K. Doibale. Study on item and test analysis of multiple choice questions amongst undergraduate medical students. International Journal of Community Medicine and Public Health 2017; 4(5): 1562-5