Analysis of multiple choice questions (MCQ): important part of assessment of medical students

Mehta M[1], Banode S[2], Adwal S[3]

[1]Dr. Maulin Mehta, Assistant Professor, [2]Dr. Siddharth Banode, Assistant Professor, [3]Dr. Sandeep Adwal, Assistant Professor, All are affiliated with Department of Pharmacology, R.D. Gardi Medical College, Ujjain, MP, India

Address for Correspondence: Dr. Maulin Mehta, Assistant Professor, Department of pharmacology, R.D. Gardi Medical College, Ujjain. E-mail: maulin_mehta@rediffmail.com

……

Abstract

Background: Assessment influences students learning process that is why analysis of assessment allows us to conduct it properly and accurately. Tarrant M et al. and few others had done similar studies in past. Further studies are required to support the same and continuing awareness among teachers about better assessment of students. **Methods:** The observational, non-interventional and prospective study was carried out to analyse 100 MCQs used for assessment of 2nd MBBS students. All MCQs were having single stem with four options, one being correct and other three distractors. Each MCQ was analyzed with three tools that were Difficulty Index (DIF I), Discrimination Index (DI) and Distracter Efficiency (DE). Chi square test was used for statistical analysis. **Results:** Total 74 out of 100 MCQs (74%) were recommended (30-70%) according to DIF I. According to DI, total 34 out of 100 MCQs were good (0.25-0.35) and 30 MCQs were excellent (>0.35). There were 18 non-functional (6%) distractors out of total 300. In none of the MCQs, all three were poor distractor. Association between difficulty index and discrimination index was statistically significant according chi-square test. (p value< 0.05) **Conclusions:** Properly constructed MCQs, according to these analysis tools, are best for student's assessment.

Keywords: Difficulty index (DIF I), Discrimination index (DI), Distractor efficiency (DE), MCQs (multiple Choice Questions), Assessment.

………..

Introduction

Proper assessment of students is very important when the learning goals involve the acquisition of skills that can be demonstrated through action in medical field. It is an essential part of the learning process in medical education. For students, assessment is a dominant motivator to direct and drive their learning [1]. Students are inclined to gain a surface approach when assessment emphasis on recall of factual knowledge. While students are more likely to adopt a deep approach if assessment demands higher levels of cognitive abilities [2]. So it has been seen repeatedly that one of the most important factor influencing students choice of learning approach is the way how assessment is being conducted [3-5]. Different methods of assessment namely Multiple Choice Questions (MCQs), Short Essay Questions (SEQs), Objective Structured Practical Examination(OSPE) and VIVA are commonly used to assess medical knowledge in undergraduate medical

education. Multiple choice questions (MCQs) are the most frequently used type of tests for assessment as it creates emphasis on knowledge and higher cognitive abilities in students. Moreover, MCQs are appropriate tool for measuring application and analysis [6]. MCQs are being used increasingly also due to their higher reliability, validity, and ease of scoring [7, 8].

As assessment influences the learning process of students, proper analysis of assessment allows us to conduct it accurately. It enables us to identify different qualities of MCQs based on different analyzing tools. Tarrant M et al., Hingorjo MR et al. and few others had done similar studies in the past which analyzed the validity of MCQs with the help of tools for proper assessment [9, 10]. Tools used in these studies were difficulty index (DIF I) also denoted by FV (facility value) or *P*-value, Discrimination Index (DI), and Distractor Efficiency (DE), in framing improved MCQs

[9, 10, 11]. Further studies are required to support the same which will continue to create awareness among teachers regarding the better assessment of students.

Tools such as, Difficulty index (DIF I, p-value), also called ease index, helps us to identify the percentage of students who correctly answered the item. It ranges from 0 - 100%. The higher percentage of score reflects that item is easier for students. Whereas, Discrimination index (DI) describes the ability of an item to distinguish between high and low scoring students. It ranges between 0 and 1. The higher score reflects the excellent ability of item to discriminate between high and low performing students. Analysis of distractor is another important part of item analysis. Distractor efficiency is one such tool that tells whether the distractors in item (MCQ) was well constructed or failed to perform its purpose in distracting students from selecting correct answer. Any distractor that has been selected by less than 5% of the students is considered to be a non-functioning distractor (NFD) [10].

Hence the present study is done with the objective to analyze MCQs (Item) with valid tools like difficulty index (DIF I) also denoted by FV (facility value) or *P*-value, Discrimination Index (DI), and Distractor Efficiency (DE), in framing improved MCQs for further assessment [11].

Material and Method

The proposed study was an observational, non-interventional and prospective in nature. Total 120 students of 2nd MBBS, after completion of cardiovascular system, in pharmacology were appeared in an internal assessment test. Internal assessment test was comprised of 100 "single response type" MCQs of 100 marks. All MCQs were have single stem with four options/responses including, one being correct answer and other three incorrect alternatives (distracter). Each correct response was awarded 1 mark. The study was conducted after getting approval from institutional ethics committee.

All MCQ answer sheets were collected from students and Each MCQ was analyzed with three tools that is Difficulty Index (DIF I), Discrimination Index (DI) and Distracter Efficiency (DE). Data obtained was entered in MS Excel 2007 and analyzed. Scores of students were entered in descending order and whole group was divided in three groups, upper 1/3 (higher ability group-HAG), middle 1/3 and lower 1/3(lower ability group). Middle group did not participate in the study. Total 100 MCQs were analyzed with various indices like DIF I, DI, and DE with following formulas: [10, 11, 12].

1. DIF I or P value (difficulty index)
    $= [(H + L)/N] \times 100$
    P value - 0 to 100%; where <30% = too difficult, 30%-70%= recommended, >70%= too easy [11]

2. DI (Discrimination Index)
    $= 2 \times [(H - L)/N]$
    DI, 0 to 1; <0.15= Discard, <0.25= Poor, 0.25-0.35= Good, >0.35= Excellent
    Higher the value of DI, item is more able to discriminate between students of higher and lower abilities. DI of 1 is ideal as it refers to an item which perfectly discriminates between students of lower and higher abilities [10]

3. DE (Distracter Efficiency)
    $= M/N \times 100$
    <5% = poor (NFD) [12]

    Where,
    N - Total number of students in both upper 1/3 and lower 1/3 groups
    H - Number of students answering the item correctly in HAG
    L - Number of students answering the item correctly in LAG, respectively.
    M – Number of students (from both groups) who choose that particular distracter

Results

Total 100 MCQs were analyzed with three different indices that is Difficulty Index (DIF I), Discrimination Index (DI) and Distracter Efficiency (DE).

Difficulty Index (DIF I): Result of DIF I of 100 MCQs was showing that 74 out of 100 MCQs were "recommended" (30-70%) and rests of the MCQs were "too easy" or "too difficult" according to Difficulty Index. [Table I] **Distractor Efficiency (DE):** In case of DE criteria, there were 18 non-functional (poor) distractors in total 100 MCQs. Out of these 100 MCQs, none of the MCQ contained all the three non-functional distractors. [Figure II]

Table I: Difficulty index (DIF I)

Difficulty Index (DI)	No. of MCQs
Too difficult (<30%)	20
Recommended (30-70%)	74
Too easy (>70%)	06

Discrimination Index (DI): On analyzing all the MCQs by DI, 34out of 100 MCQs were "good" (0.25-0.35) and 30 MCQs in "excellent" criteria, while rests of the MCQs were "discard/poor" according to DI criteria. [Table II]
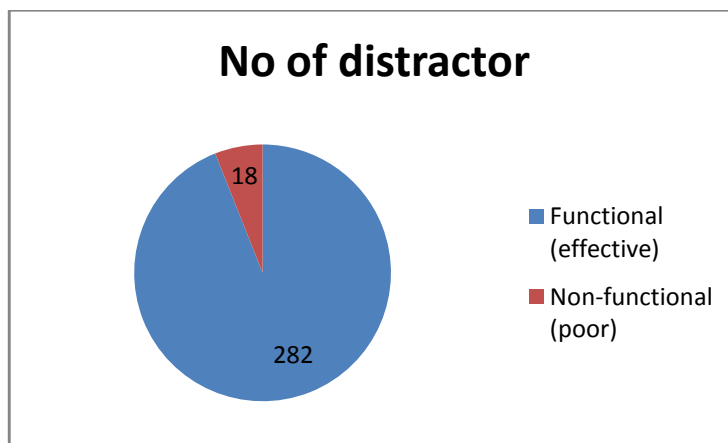
Table II: Discrimination Index (DI)

Discrimination Index (DI)	No. of MCQs
Discard (<0.15)	20
Poor (<0.25)	16
Good (0.25-0.35)	34
Excellent (>0.35)	30

Figure I: Association between difficulty and discrimination index
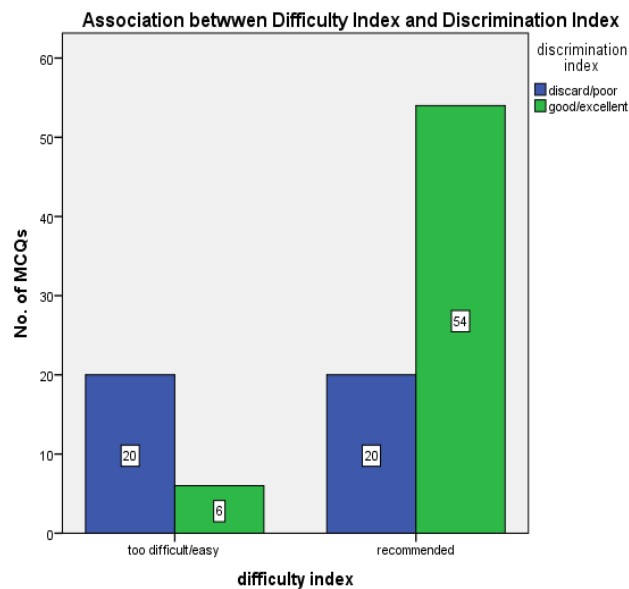
Figure II: Distractor efficacy

Discussion

Many changes have been made so far to medical curriculum to assess students to produce more reflective, self-directed medical practitioners. Students may combine 3 broad approaches to learning and studying: deep, surface, and strategic [13]. The surface approach includes memorizing without understanding and restriction to the syllabus, not recognizing the wider context. It is often driven by fear of failure and lack of purpose. While students with a deep approach seek meaning in material, are interested in ideas, relate new ideas to previous knowledge and use evidence critically.

The MCQ format allows teachers to efficiently assess deep approach of large numbers of candidates and test a wide range of content [14, 15]. If properly constructed, MCQs are able to test higher levels of cognitive reasoning and can accurately discriminate between high- and low-achieving students [14, 16]. It is widely accepted that well-constructed MCQ items are time consuming and difficult to write [17]. So it is essential to analyse MCQ for its appropriateness.

One of the tools for analysing MCQs is Difficulty Index (DIF I). It categorized MCQ as "too difficult", "recommended" or "too easy". In our study, we observed that 74% out of total MCQs were "recommended" according to DIF I. Instructional Assessments Resources (IAR) insinuates the usage of easy question as warm up questions in assessing student mastery, but it will not test higher level of cognitive reasoning. On the other side, "too difficult" MCQs will only check the ability of good students. This will mainly affect the poor students and will benefit good students. So by using "too difficult" MCQs, we will not be able to discriminate between good and poor performing students. [18] So the MCQs ought to be reconsidered in terms of language and content appropriateness to make it "recommended" according to Difficulty Index (DIF I). MCQs in "recommended" category are appropriate in all sense for both higher (good) and lower (poor) ability groups. Hence in our study, remaining 26% of MCQs required changes in language or content to make it "recommended" according to DIF I.

Discrimination index (DI) is also very important tool for analysis of MCQs and discriminate students between higher and lower ability groups, which is essential part of better assessment. In our study, 64% of total MCQs were "good" or "excellent" discriminator according to DI. We cannot assess all the medical students by using common types of MCQs because all the students are not of equal caliber. MCQs with higher Discrimination Index (>0.35) will be an "excellent" discriminators. MCQs with discrimination index < 0.25 are not going to help teachers to discriminate their students between two groups. It is seen that moderately difficult ("recommended") questions had better discriminating power than "too difficult" or "too easy" questions. So Discrimination index is always associated with difficulty index [15] In our study, total 54 "recommended" MCQs according to difficulty index were also "good/excellent" MCQs as per discrimination index. Chi-square test p value 0.0001 (p value <0.05) was showing that association was statistically significant between Difficulty Index and Discrimination Index. This confirms one of the basic principles of item response theory, which postulates that the questions that are at the same level of the average candidate's ability are the most effective to assess/discriminate them [11,19].

One aspect where many MCQs fail is in having effective distractors. "Non-functioning" or "poor" distractors are options that are selected infrequently (<5%) by examinees or otherwise do not perform as expected. Teachers often spend a great deal of time in constructing the stem and much less time on developing effective distractor to the correct answer. High quality MCQs also need the options to be well written. [18] In a classroom setting where test items are designed to measure educational outcomes, distractors must perform acceptably and each distractor should be based on a common misconception about the correct answer [19]. With non-functioning distractors the questions become "too easy" (DIF I) and "poor discriminating" (DI). In our study, less non-functioning distractors (6%) correlate with higher percentage of "recommended" (74%, difficulty index) and "good/excellent" (64%, discrimination index) MCQs. These non-functioning distractors should be removed and replaced with functioning distractors for future assessment of other students.

In another study regarding number of distractors in single MCQ, Haladyna et al. found that approximately two-thirds of all four-option items they reviewed had only one or two functioning distractors and none of the five-option items had four functioning distractors [20].

Because it is often, too difficult for teachers to develop three or more equally plausible distractors, additional distractors are often added as "fillers." [16, 21] More numbers of distractors are not necessarily better while making distractors – the key is the quality of the distractors, not the number [18]. A meta-analysis of 80 years of research on the number of options in MCQs also concluded that three options are optimal for MCQs in most settings [22]. In our study, we under took four-option items for assessment. But in future, if we undergo such study we would try for three-option items, which will go in flow with the previous meta-analysis.

Assessment of MCQs by these indices highlights the importance of assessment tools for the benefits of both the students and the teachers.[23] Results gathered from this study highlighted the importance of item analysis by using Difficulty Index, Discrimination Index and Distractor Efficiency. Items having average difficulty and high discrimination power with functioning distractors should be incorporated into future tests to improve the assessment. In our study, less numbers of MCQs (items) were limitation. This can be omitted by taking large sample size for more significant result.

Conclusion

Properly constructed MCQs by using these analysis tools are best for students' assessment. More "recommended" MCQs with all "functional distractors" are good discriminators between high and low ability group of students. All the MCQs should be made "recommended", "excellent" and with "functional" distractor for better assessment of students. Modified MCQs, according to analysis tools, should be stored in computer as MCQs bank in department and can be reused for assessment of other students in future also. Using the analysis tools described in this article are a necessary condition for constructing adequate multiple choice questions. The distractors used in MCQs should be functional without exception. If this is the case then these items will be able to discriminate adequately between high and low ability group of students. MCQs modified with the help of tools described here can be reused when new group of students will be assessed.

References

2. Scouller K. The influence of assessment method on students' learning approaches: multiple-choice question examination versus assignment essay. Higher Educ. 1998; 35:453–72.

3. Trigwell K, Prosser M. Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. Higher Educ. 1991; 22:251–66.

4. Biggs, J. Teaching for Quality Learning at University. The Society for Research into Higher Education & Open University Press, Philadelphia, 2003.

5. Reid WA, Duvall E, Evans P. Relationship between assessment results and approaches to learning and studying in Year Two medical students. Med Educ. 2007 Aug;41(8):754-62.

6. Abdel-Hameed AA, Al-Faris EA, Alorainy IA, Al-Rukban MO. The criteria and analysis of good multiple choice questions in a health professional setting. Saudi Med J. 2005 Oct;26(10):1505-10.

7. Case S, Swanson D. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia: National Board of Medical Examiners, 2003.

8. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurse Educ. 2012 May-Jun;37(3):98-104. doi: 10.1097/NNE.0b013e31825041d0.

9. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Med Educ. 2009 Jul 7;9:40. doi: 10.1186/1472-6920-9-40

10. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012 Feb;62(2):142-7.

11. Singh T, Gupta P, Singh D. Principles of Medical Education. 3rd ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd. Test and item analysis. 2009; 70

12. Cizek GJ, O'Day DM: Further investigations of nonfunctioning options in multiple choice test items. Educ psycho meas. 1994, 54(4):861-72.

13. Entwistle N. Styles of Learning and Teaching. London: David Fulton Publishers. 1997; 77–107.

14. Downing SM: Assessment of knowledge with written test forms. In International handbook of research in medical education Volume II. Edited by: Norman GR, Van der Vleuten C, Newble DI. Dorcrecht: Kluwer Academic Publishers. 2002; 647-72.

15. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach. 2004 Dec;26(8):709-12.

16. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ. 2004 Sep;38(9):974-9.

17. Farley JK. The multiple-choice test: writing the questions. Nurse Educ. 1989 Nov-Dec;14(6):10-2, 39.

18. Haladyna TM, Downing SM: Validity of taxonomy of multiple choice item-writing rules. ApplMeasEduc. 1989; 2(1):51-78.

19. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore. 2006 Feb;35(2):67-71.

20. Haladyna TM, Downing SM: How many options is enough for a multiple-choice test item? Educ Psychol Meas. 1993; 53(4): 999-1010.

21. Crehan KD, Haladyna TM, Brewer BW: Use of an inclusive option and the optimal number of options for multiple-choice items. Educ Psychol Meas. 1993; 53(1):241-7.

22. Rodriguez MC: Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educ Meas Issues Pract. 2005; 24(2):3-13.

23. Pellegrino J, Chudowsky N, Glaser R, editors. Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC: National Academic Press 2001.

………………………….
How to cite this article?

…….