Original article

Item analysis: An unmatched tool for validating MCQs in Medical Education

Dr. Shikha Saxena[1], Dr. P. C. Srivastava[2], Dr. Ayaz k. Mallick[3], Dr. H. S. Joshi[4], Dr. Biswajit das[5]

1- MD, DGO,Assistant Professor,Department of Biochemistry Rohilkhand Medical College & Hospital Bareilly - 243006, U.P., INDIA.

2- MD,Professor,Department of Forensic Medicine,Rohilkhand Medical College & Hospital,Bareilly, U.P.

3- MD ,Associate Professor,Department of Biochemistry ,Rohilkhand Medical College & Hospital Bareilly , U.P.

4- Professor & Head,Dept. of Community Medicine,Rohilkhand Medical College & Hospital ,Bareilly, U.P.

5- MD ,Professor & Head,Department of Biochemistry,Rohilkhand Medical College & Hospital Bareilly, U.P.

Corresponding author: DR. Shikha saxena, MD, DGO

Abstract:

Background: Assessment of learning process and evaluation of assessment tool is an important component of a teaching-learning curriculum. Multiple choice questions (MCQs) are preferred over other tools of assessment. The unique advantage of multiple-choice questions is that one can get statistical information by analysing the p-erformance of MCQs, known as an item analysis. The aim and objectives of the study was to analyse and assess the quality of MCQs by item analysis for creating viable questions.

Methods: Assessment of 80 first professional MBBS students was done by 30 MCQs after completing intera-ctive didactic lectures. Post validation item analysis was done for Facility value (FV), Discrimination index (D-I), & Distractor efficiency (DE).

Results: Facility value or Difficulty index of 70.00% items was in the acceptable range, 3.33% item was too ea-sy and 26.67% items was too difficult. Discrimination index of 26.67% items was within recommended DI val-ue, 53.33% items was within acceptable range, and 20% items fell within discarded category. Items with all fu-nctional distractors constituted 86.67% whereas only 13.33% items were with single non-functional distractor. Considering FV, DI and DE together, 23.33% items were validated for MCQ Bank, 56.67% items would be re-validated after revision and modification, and 20% items were discarded.

Conclusion: Items having average difficulty and high discriminating power with functional distractors should be incorporated into future tests to improve the quality of assessment development and its review.

Key Words: Item Analysis, Medical Education

Keynotes: Item Analysis, Feedback.

Introduction:

Assessment of learning process and evaluation of assessment tool is an important component of a teaching learning curriculum. A significant applic-ation of assessment is the continuous monitoring of learning activities for giving a feedback to students and teachers. Today multiple choice questions (M-CQs) is one of the most comm.-only used tools for assessing the knowledge of medical students.[1] One of the unique advantages of multiple-choice questions is that one can get statistical information by analysing the performa-nce of MCQs about how well one's questions are working, known as an item analysis.[2] It is said that MCQs emphasize recall of factual infor-mation rather than concept-ual understanding and interp-retation of concepts

.[3]Properly construct-ed MCQs canas-sess higher cognitive processing of Bloom's taxonomy such as interpretation, synthesis and application.[4,5] Multiple choice questions is preferred over other tools of assessment because of its objectivity in assessent, com-parability in different settings, wide coverage of subject, and minimization of assessor's bias.Designing good MCQs is a complex, challenging and time consuming process. Having constructed and ass-essed, MCQs need to be tested for the standard or quality. Item analysis examines the student responses to in-dividual test items (MCQs) to assess the quality of those items and test as a whole.[6]It is a valuable yet relatively simple procedure performed after the examination that provides information regarding the reliability and validity of a test.[7]Thus item analysis assesses the assessment tool for the benefit of both student and teacher. To as-sess the students by MCQs,we either frame or take help of old MCQsfrom question papers/text books whose va-lidation status is un-knwn. Therefore, we conducted this study to frame the MCQs which was used as assess-ent tool of the stud-ents, later on to analyze the quality of MCQs, to improve the items that needed modification and for creatin-g a viable question bank for subsequent use.

Aim &Objectives:

The aim and objectives of the study was to analyse and assess the quality of multiple choice questions by item analysisfor creating viable questions in Biochemistry for future use.

Material & Methods:

The project was conducted in the department of Biochemistry, Rohilkhand Medical College, Bareilly, Uttar Pradesh. Ethical approval for the same was taken from the Institutional Ethical Committee. Informed and un-derstood consent was taken from the subjects prior to participation in the study. Assessment of first professional MBBS students was done by MCQs after completing interactive didactic lectures. Prior notification was done regarding date, time and topics for the MCQ test. Out of hundred, ninety three students appeared for the test co-mprising of thirty items with single correct response each carrying one mark. The time allotted was 40 minutes and there was no negative marking. Post validation of the paper was done by item analysis. Each item was ana-lysed for Facility value (FV in %), Discrimination index (DI), & Distractor efficiency (DE in %) and mean and standard deviation was calculated using Microsoft Office Excel 2007.

The scores of all the students were arranged in order of merit. The students with score ≥14 were considered as High Ability Group (HAG) and those with ≤13 were placed in Lower Ability Group (LAG). To make equal nu-mber of students in each group (40 each), thirteen students having score 13 were dropped out. However, only o-ne student with score 13 was retained in LAG and one student with score 14 was dropped out from HAG. An it-em (a MCQ) contains a stem and four options including one correct (key) and three incorrect (distractors) alter-natives. Each item was analysed for:

- **1. Facility value (FV) or Difficulty Index (Dif I):** It is the number of students in the group answering a que-stion right. It is calculated using the formula:**FV = (HAG + LAG) × 100 / N**

  HAG = number of students answering the item correctly in the high ability group, LAG = number of students answering the item correctly in the low ability group, N = total number of students in the two groups.Facility value is the measure of how easy or how difficult a question is. Higher the FV, easier is the question.
- **2. Discrimination index (DI):** This index indicates the ability of a question to discriminate between a

www.ijbamr.com P ISSN: 2250-284X , E ISSN : 2250-2858

higher and a lower ability student. This is calculated by using the formula:

DI = 2 X (HAG - LAG)/N where the symbols HAG, LAG and N represent the same values as mentioned above.

- **3. Distractor Efficiency (DE):** On the basis of number of Non Functional Distractors (NFDs) in an item, DE ranges from 0 to 100%. An NFD in an item is the option, other than the key selected by less than 5% of students and functional or effective distractor is the option selected by 5% or more students.[8] If an item contains three or two or one or nil NFDs then DE would be 0, 33.3%, 66.6% and 100% respectively.[9] In general, items with a FV value between 30 - 70% are considered as acceptable. Items with FV value <30% (too difficult) and >70% (too easy) are not acceptable and need modification. The Discrimination index (DI) is a measure of the item to discr-iminate between students of higher and lower abilities and ranges between 0 and 1. In general, the recommende-d DI value is >0.25 and DI value 0.15 - 0.25 is acceptable with revision whereas DI value <0.15 is discarded.[8]

Results:

Facility value orDifficulty index of 21 (70.00%) items was in the acceptable range, 1 (3.33%) item was too easy and 8 (26.67%) items was too difficult. The FV ranged between 8.75% - 82.5% with the mean 43.42 ± 18.68. [Fig. 1 & Table - 1]Discrimination index of 8 (26.67%) items was within recommended DI value, 16 (53.33%) items was within acceptable range, and 6 (20%) items was poor and fell within discarded category. Among discarded category items, two were having negative DI value of -0.03. The mean DI calculated was 0.21 ± 0.11. [Fig. 2 & Table -1]

A total of thirty items had 90 distractors. Amongst these, 26 (86.67%) items were with all functional distractors whereas only 4 (13.33%) items were with single non-functional distractor. The mean DE was 95.55 ± 11.55. On the basis of non-functional distractors, distractor efficiency of each item was assessed. [Fig. 3 & Table -1]

Considering facility value (FV), discrimination index (DI) and distractor efficiency (DE) together, 7 (23.33%) items were validated, 17 (56.67%) items would be re-validated after revision and modification, and 6 (20%) items were discarded.

Table 1: Comparison of FV, DI, and DE of the MCQs (items)

Parameter	FV	DI	DE
Range	08.75% – 82.50%	(-0.03) – 0.53	66.60% – 100%
Mean ± S.D.	43.42 ± 18.68	0.21 ± 0.11	95.55 ± 11.55

Note: FV = Facility Value, DI = Discrimination Index, DE = Distractor Efficiency, MCQs = Multiple Choice Questions, S.D. = Standard Deviation.

www.ijbamr.com P ISSN: 2250-284X , E ISSN : 2250-2858

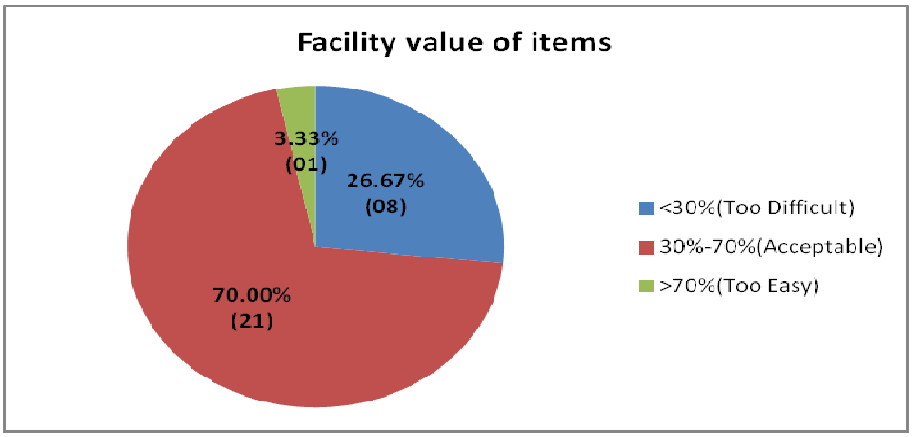

Fig. 1: Facility value of Items

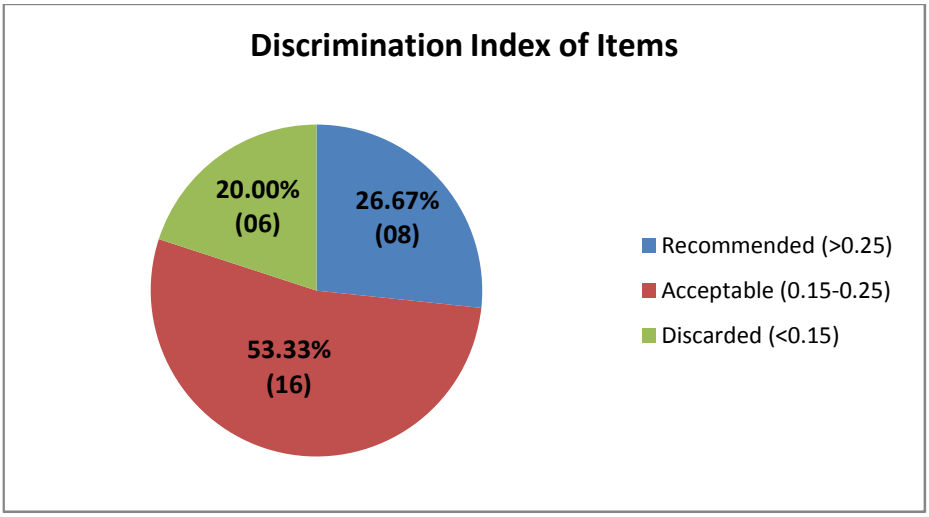

Fig. 2: Discrimination index of items

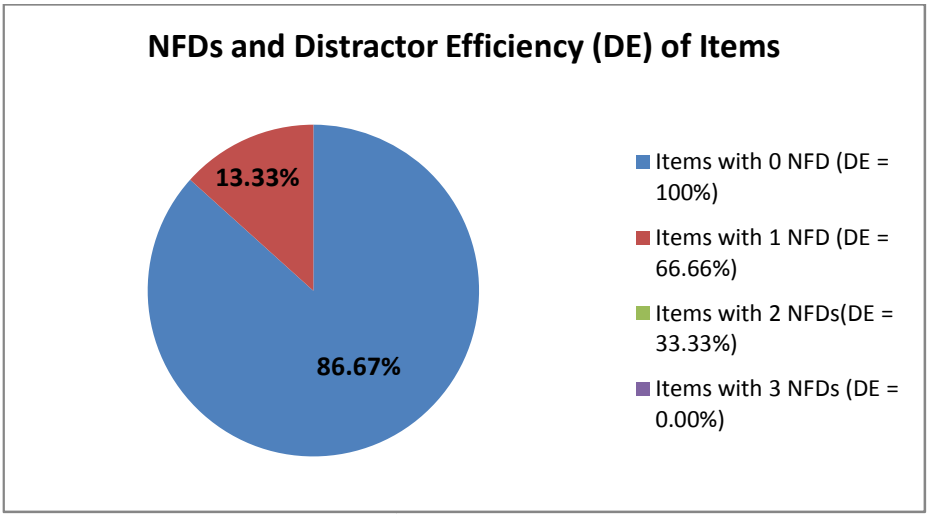

Fig. 3: Non-Functional Distractors and Distractor Efficiency of the items

www.ijbamr.com P ISSN: 2250-284X , E ISSN : 2250-2858

Discussion:

Any assessment whether formative or summative, has intense effect on learning and is an important variable in directing the learners in a meticulous way.[10] Single correct response type MCQ is an efficient tool for evaluate-on. [11] However, this efficiency solely rests upon the quality of MCQ which is best assessed by the analysis of it-em and test as a whole, together referred as item analysis. It also helps to identify the subject content which lac-ks understanding and need greater emphasis and clarity, by improving or changing the methodology of teachi-ng. Poor items can be modified or removed from the pool of questions.

In previous studies, Gajjar et al[9] and Pande et al[12] have proposed the mean of Dif I (F.V.) as 39.4 ± 21.4% and 52.53 ± 20.59respectively**.** Karelia et al[13] showed a range of F.V. between 47.17 ± 19.77 to 58.08 ± 19.33 in their study over a period of five years.They also showed 61% items in acceptable range (30-70%), 24% items in too easy (>70%) and 15 % items in too difficult range (<30%). Another study[12] showed that 62% items had F.V of 30-70%, 23 % were too easy and 15% were too difficult.Patel and Mahajan[14] showed 80% of items in the acceptable range and 20% in the unacceptable range.Mehta and Mokhasi[1] in their study noted a mean of Dif I/F.V as 63.06 ± 18.95. Majority of the items (62%) were in the acceptable range whereas 32% items were too easy and only 6% items were too difficult. Mehta and Mokhasi[15]in another study noted 70% acceptable, 26% too easy and 4% too difficult out of a total of 50 items. Mean difficulty index (Dif I) in a study conducted am-ong medical students of Ahm-edabad[9] was 39.4 ± 21.4% and 48% items were considered within the acceptable range. In the present study we calculated the mean FV (Dif I) of 43.42 ± 18.68 which was within the acceptable range. Our acceptable percentage of items (70%) corresponded with those of previous studies[1,12,13,14] whose val-ues ranged from 61 to 80%.

The Dif I and DI are often reciprocally related as higher the Dif I, lower is the difficulty of the question. Que-stions having high FV (easier questions) discriminate poorly; conversely questions with a low FV are conside-red to be good discriminators.[16]But our study failed to prove this reciprocal relationship between Dif I & DI. In the present study, the mean DI was 0.21 ± 0.11. Items with DI >0.25 were 8 (26.67%), DI between 0.15 and 0.25 were 16 (53.33%) and DI <0.15 were 6 (20%). There were 2 items with negative DI. Earlier study[11] has revealed 40% items with DI >0.35, 42% with DI between 0.2 and 0.34 and 18% with DI <0.20. In another study[1], the mean DI was 0.33 ± 0.18. Items with DI >0.35 were (52%), DI between 0.2 and 0.34 were (18%) and DI <0.2 were (30%). Mean DI in a recent study[9] was 0.14 ± 0.19 less than the acceptable cut off point of 0.15.[6] It was so because 10/50 items had DI less than zero (negative DI).[9] Another study[11] had the mean DI of 0.36 ± 0.17 and had only 2/50 items with negative DI. Negative DI indicates that the students of lower ability answer more correctly than those with higher ability. This is probably due to complex nature of item, wrong key, ambiguous framing of question[17] or generalized poor preparation of students, making it possible for lower ability group of students to select correct response by guess without any real understanding about the question related topic, while a good student takes a harder path to solve and ends up to be wrong.Constructing the plausible distractors in an item is as important as the key. Although the correct answer must be truly correct, it is just as important that the distrac-tors be incorrect. Properly constructed distractors should be able to distract LAG students while HA-G students should minimally opt for them. More NFD in an item increa-ses DIF I (makes item easy)

www.ijbamr.com P ISSN: 2250-284X , E ISSN : 2250-2858

and reduces DE, conversely item with more functioning distractors decreases DIF I (makes item difficult) and increases DE. Higher the DE more difficult the question and vice versa, which ultimately relies on presence/absence of NFDs in an item.

In the present study with thirty MCQs, having 90 distractors, 04 (4.44%) were found to be NFDs, 86 (95.55%) were functional distracters. Mean DE in present study was 95.55 ± 11.55. We found no items with 2 or more NFDs. In a study conducted on 514 items and 1542 distractors, 35.1% were NFDs, 52.2% were functional distr-actors and 10.2% were not chosen by any student.[18]Another review of functioning distractors in 477 items on f-our MCQ assessments showed 38% itemshad NFDs and items with three functional distractors ranged from on-ly 1.1 to 8.4%.[19]In another study[1] with fifty MCQs, having 150 distractors, 53 (35.33%) were found to be NF-Ds, 28 (18.66%) were functional distractors and 69 (46.01%) distractors had nil response. Items with DE 66.6% were 18 (54.4%), items with DE 33.3% were 9 (27.27%) and items with DE as 0 were 6 (18.18%). The remain-ing 17 items with three functional distractors had DE as 100%. Gajjar et al[9] have noted 133/150 (89.6%) functi-onal distractors, and 17 (11.4%) were NFDs present in 15 items (13 had 1 and two had 2) with DE varying bet-ween 33 and 66%. Remaining 35 items had no NFDs with their DE being 100%. Considering the FV, DI & DE together, we validated 7 (23.33%) as ideal items in our study similar to those of other researchers[1,13] who validated 12 (24%) and 15(30%) items respectively.

Conclusions:

This study inferred that items having average difficulty and high discriminating power with functional distr-actors should be incorporated into future tests to improve the quality of assessment development and its review. Other benefits of Item analysis are:

1. Teacher would be able to prepare MCQs, analyse and validate for future use.
2. Lower ability students and their learning difficulties would be identified which could be corrected by giv-ing feedback, counselling and/or modifying learning methods.
3. Students would be put into the process of self-motivated deeper study directed towards their desired learn-ing outcome.
4. The teachers would also get a feedback on the efficacy of their teaching, for improvement of teaching skills in the future.

References:

1. 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.
2. "Open Professionals Education Network (OPEN) Resource Document Writing Multiple-Choice Questions: Item Analysis" Available at http://static1.squarespace.com/static/5330c47be4b03ea35b2645a8/t/5481ef08e4b0a7209a9cd49c/1417801480204/MC+questions+-+Item+analysis+final.pdf (accessed on 15 May 2015)
3. Srivastava A, Dhar A, Agarwal CS. Why MCQ? Indian Journal of Surgery. 2004; 66:246-48.
4. Carneson J, Delpierre G, Masters K. Designing and managing MCQs. Appendix C: MCQs and Blooms tax-onomy (Online) 2011 (Cited 2011 Feb2). Available at http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html. (Accessed on 10 Aug 2015)
5. Case SM, Swanson DB. Extended matching items: a practical alternative to free response questions. Tech Learn med 1993;5:107-15.

6. Singh T, Gupta P, Singh D. Test and Item analysis. In: Principles of Medical Education. 3rd ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd; 2009;70-77.
7. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for u-se in nursing research and education. Collegian 2005;12:19-24.
8. Singh T, Gupta P, Singh D. Test and Item Analysis. In: Principles of Medical Education. Fourth ed. New Delhi Jaypee Brothers Medical Publishers (P) Ltd; 2013:108-13.
9. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality Multiple Choice Questions (MCQs) from an assessment of medical students of Ahemdabad, Gujarat. Indian journal of Community M-edicine. 2014;39:17-20.
10. Sarin YK, Khurana M, Natu MV, Thomas AG, Singh T. Item analysis of published MCQs. Indian Pediatr 1998;35:1103-5.
11. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and dist-racter efficiency. J Pak Med Assoc 2012;62:142-47.
12. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discriminate-ion indices of MCQs in formative exam in Physiology. South East Asian Journal of Medical Education 20-13;7(1):45-50. Available at http://seajme.md.chula.ac.th/articleVol7No1/OR6_Sushma.pdf (accessed on 02 Jan 2016)
13. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship betw-een them in four response type multiple choice questions of pharmacology summative tests of year II MB-BS students. Ie JSME. 2013;6:41-46.
14. Patel KA, Mahajan NR. Itemized Analysis of Questions of Multiple Choice Question (MCQ) Exam. Inter-national Journal of Scientific Research 2013;2:279-80.
15. Mehta G, Mokhasi V. Efficacy of multiple choice questions as a tool for formative assessment. Internatio-nal Journal of Development Researc. 2015;5(01):2941-44.
16. Carroll RG. Evaluation of Vignette-type examination items for testing medical physiology. Am J Physiol 1993;264:S11-5.
17. Scantron Guides-Item Analysis,adapted from Michigan State Universitywebsite and Barbara gross devils tools for teaching.Available rom: www.freepdfdb.com/pdf/item-analysis-scantron [Last cited on 2013Apr 13].
18. Tarrant M, WareJ, Mohammed AM. An assessment of functioning and nonfunctioning distractors in mult-iple choice questions: a descriptive analysis. BMC Medical Education 2009;9:40.
19. Haladyna TM, Downing SM. How many options is enough for a multiple choice test item? Educ Psychol Meas. 1993;53:999-1010.