International Journal of Community Medicine and Public Health
Ingale AS et al. Int J Community Med Public Health. 2017 May;4(5):1562-1565
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: http://dx.doi.org/10.18203/2394-6040.ijcmph20171764

Original Research Article

Study on item and test analysis of multiple choice questions amongst undergraduate medical students

Abhijeet S. Ingale[1]*, Purushottam A. Giri[2], Mohan K. Doibale[3]

Department of Community Medicine, [1]Assistant Professor, [2]Professor, JIIU's Indian Institute of Medical Science and Research Medical College, Badnapur, Jalna, Maharashtra, India
[3]Department of Community Medicine, Professor and Head, Govt. Medical College, Aurangabad, Maharashtra, India

Received: 10 March 2017
Accepted: 03 April 2017

***Correspondence:**
Dr. Abhijeet S. Ingale,
E-mail: iabhijeet@yahoo.com

ABSTRACT

Background: Item analysis is the process of collecting, summarizing and using information from students' response to assess the quality of test items. However it is said that MCQs emphasize recall of factual information rather than conceptual understanding and interpretation of concepts. There is more to writing good MCQs than writing good questions. The objectives of the study was to assess the item and test quality of multiple choice questions and to deal with the learning difficulties of students, identify the low achievers in the test.
Methods: The hundred MBBS students from Government medical college were examined. A test comprising of thirty MCQs was administered. All items were analysed for Difficulty Index, Discrimination Index and Distractor Efficiency. Data entered in MS Excel 2007 and SPSS 21 analysed with statistical test of significance.
Results: Majority 80% items difficulty index is within acceptable range. 63% items showed excellent discrimination Index. Distractor efficiency was overall satisfactory.
Conclusions: Multiple choice questions with average difficulty and also having high discriminating power with good distracter efficiency should be incorporated into student's examination.

Keywords: Item, Difficulty index, Discrimination index, Non functional distracters, Distracter efficiency

INTRODUCTION

Item analysis is the process of collecting, summarizing and using information from students' response to assess the quality of test items.[1,2] However it is said that MCQs emphasize recall of factual information rather than conceptual understanding and interpretation of concepts.[3] There is more to writing good MCQs than writing good questions. Properly constructed MCQs can assess higher cognitive processing of Bloom's taxonomy such as interpretation, synthesis and application of knowledge, instead of just testing recall of isolated facts.[4,5]

A multiple-choice question (MCQ) consists of a stem with a question line at its end or underneath it, followed by a number of options. One of the options is the correct or best response known as the key, while the others are described as distracters.[6] An essential characteristic of distracters is that all options shall present plausible answers and if possible none shall be incorrect.[7] Function of a distracter is to attract students who do not know the correct answer while students who know the correct answer ignore them. Multiple choice questions (MCQ'S) are used as an objective and reliable tool to evaluate learning performance of students. It is also a preferred tool for selection of students for a given course. MCQ needs to be tested for the standard or quality.[8] However the items to be used must be of good quality to serve the purpose for it is meant to be.

Objectives

- To assess the item and test quality of multiple choice questions.

- To deal with the learning difficulties of students, identify the low achievers in the test.

METHODS

A cross-sectional study was carried out in the department of Community Medicine of Govt. Medical College, Aurangabad, Maharashtra after a 'post end' assessment exam in August 2015. After a semester of didactic teaching, out of the 121 students, 100 third-year MBBS students took the MCQs test comprising of 30 questions with single best response. Similar study as a pilot project was taken in April end of 2015 year, the results of which are not included in this study and presented separately in research form. This study was a more refined version of the previous study. One mark for each question was there while no negative marking and the time allotted was half hour. Multiple answers and no response were given 0 marks. Evaluation was done out of thirty marks and 50% score was the passing mark. 30 MCQs with 120 distracters were analysed. Informed consent was sought. Pre-validation of the paper was done by team of teachers including final scrutinization by the Head of the Department. Post validation of the paper was done by item analysis by the above authors after due training. The scores of all the students were arranged in order of merit.

The upper one third students were considered as high achievers and lower third as low achievers. Difficulty index/ facility value or P value using formula P = H + L/N ×100 where H = Number of students answering the item correctly in the high achieving group L = Number of students answering the item correctly in the low achieving group N = Total number of students in the two groups (including non-responders) Discrimination index (D)/ Point Biserial correlation or d value using formula, d = H – L × 2/N where the symbols H, L, and N represent the same values as mentioned above Distractor effectiveness (DE) or functionality is based on number of non functional distracters. Values and cut-off point of various indices is based on standard textbook.[9] To avoid possible copying from neighboring student, they were administered one of three paper sets which were prepared with disorganized sequencing of questions.

Statistical analysis was done by entering the score of 100 students in order of merit in the MS Excel 2007 and SPSS 21. Data analysed with simple proportions, mean, standard deviations, pearson correlation. Reliability of the test was assessed by estimating the Kuder-Richardson 20 coefficient (KR20). The widely-accepted cut-off value of KR20 is greater than or equal to 0.7.[7,10]

RESULTS

Figure 1 showing difficulty index (p-value) of MCQ items 7% MCQ's were having difficulty index < 30, while 13% were having too difficult >70 DI. Maximum number 80% were in the acceptable range 30-70 of Diff-I.

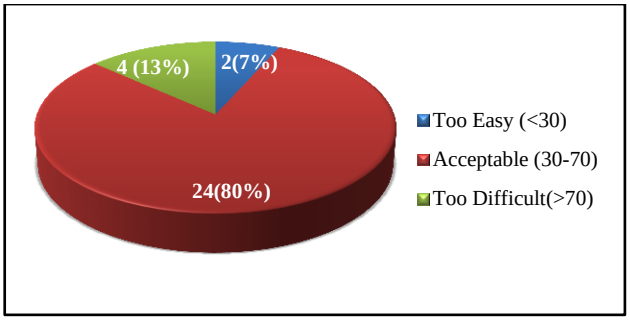

Figure 1: Difficulty index of MCQ items.

Figure 2 showing discrimination index (d value) of MCQ items. Maximum number i.e. 63% were showing excellent Discrimination index >0.35 followed by 27% with a poor i.e. <0.20 of discrimination index.

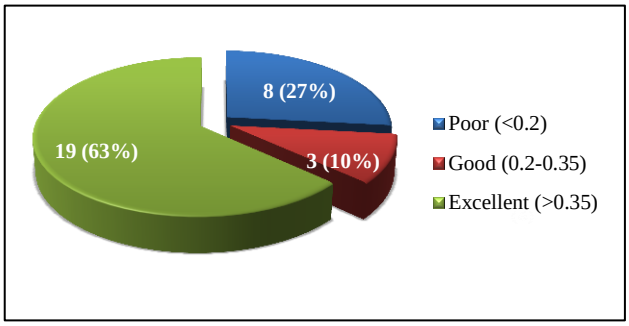

Figure 2: Discrimination index of MCQ items.

Table 1: Non-functional distracter and distracter effectiveness.

MCQ items (n=30)			
Items with 0 NFD	Items with 1 NFD	Items with 2 NFD	Items with 3 NFD
DE = 100%	DE=66.6%	DE=33.33%	DE=0%
08 (26.66%)	13(43.33%)	07 (23.33%)	02(6.66%)

NFD- Non functional distractor, DE- Distractor effectiveness.

Table 1 showing the distribution of NFDs and distractor effectiveness (DE) of each MCQ item. 26.6% MCQ's were having 100% DE with 0 NFD. 7% MCQ's were having 0% D.E with 3 NFD. While almost 66.66% MCQ's were in between the range of 33.3% to 66.6% of D.E with 1 to 2 NFD's.

Table 2 showing the comparison of Diff I, DI and DE of the MCQ items. Over all difficulty index of the items were in the range of 95.45 - 22.72 having a mean 55.10 with standard deviation of 17.28. The discrimination index of the items were in the range of 0.90 – (-0.2) having a mean 0.40 with standard deviation of 0.33. The distractor efficiency of the items was in the range of 0 – 100 having a mean 59.21 with standard deviation of 31.45 Figure 3 is showing the Scatter plot showing relationship between difficulty index (Diff I) and

discrimination index (DE) of items. The maximum discrimination (D = 0.5–0.6) was observed in acceptable range of difficulty index. i.e. Diff -I (P =30–70%). showing an inverse trend of relationship between the two indices. The discrimination index showed poor negative correlation with the difficulty index, which was found to be non-significant. (Pearson co-relation) r = -0.225, p-value= 0.232. The overall reliability of the test as measured by the above mentioned formula of Kuder Richardson 20 co-efficient (KR20) which was found to be 0.61 (The widely accepted cut off value ≥ 0.7).[7,10]

Table 2: Comparison of Diff I, DI and DE of the MCQ items.

Parameter	Diff I	DI	DE
Range	95.45 - 22.72	0.90 – (- 0.2)	0 – 100
Mean ± SD	55.10 ± 17.28	0.40 ± 0.33	59.21 ± 31.45

Diff I- Difficulty index, DI- Discrimination index, DE- Distractor effectiveness.

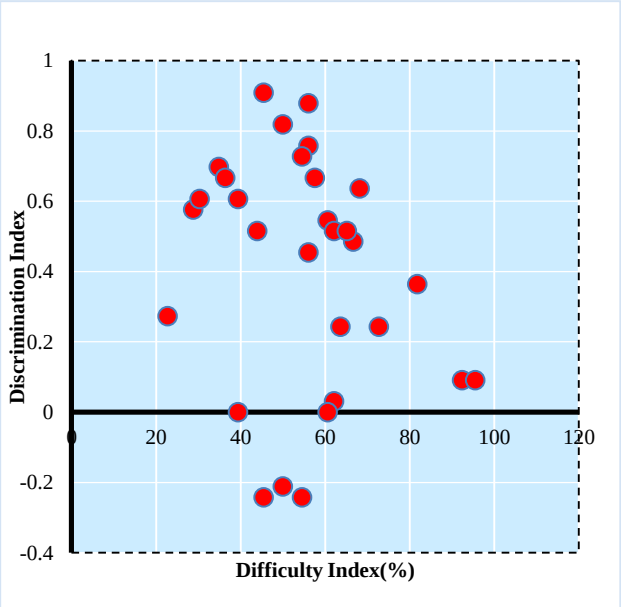

Figure 3: Scatter plot showing relationship between difficulty index (Diff I) and discrimination index (DE) of items.

According to inter item correlation matrix and overall item statistics, if we remove the item number 6,16,26 that will increase the Chronbach's alpha value to (0.66) which will be near to the acceptable value of the test i.e. 0.7 of Kuder Richardson 20 coefficient. Ultimately the chronbach's alpha will increase and increase in overall test reliability. We identified 33 low performing students. They were personally consulted by the teachers and their problems identified. All students were sorted out according to their hindrances and difficulties. They were counselled properly and teachers helped them in their areas of concern by modifying learning methods.

DISCUSSION

The study was conducted in order to assess the item and test quality of multiple choice questions, by evaluating with test indices so that a valid pool of items can be created. In view of designing a competency based test these types of studies should be frequently done. These types of studies help us to identify the low achievers in order to deal with their learning difficulties by counselling or modifying learning methods proposed accordingly. In our study the difficulty index (p-value) of MCQ items 7% MCQ's were having difficulty index <30, while 13% were having too difficult >70 DI. Maximum number 80% were in the acceptable range 30-70 of difficult index. Difficulty index of the items were in the range of 95.45 - 22.72 having a mean 55.10 with standard deviation of 17.28. Previous studies have proposed the mean of difficult index as 39.4+ 21.4%, 52.53±20.59.[11,12] Karelia, showed a range of mean±SD between 47.17±19.77 to 58.08±19.33 in a study conducted over a period of five years.[13] They also showed 61% items in acceptable range (p 30-70%), 24 % items (p >70%) and 15% items (p <30%). Other studies showed that 62% items had p value (30-70%), 23 % were too easy (p >70%) and 15% were too difficult (p <30%).[12] Patel and Mahajan showed 80% of items in the acceptable range (p 30-70%) and 20% in the unacceptable range (p >70% and <30%).[14] Our study findings were in congruence with the previous studies having a mean of difficult index as 57.92±19.58. The p value of 26 (65%) items was in acceptable range (30-70%), 10 (25%) items were easy with p value >70% and 4 (10%) items were difficult with p value <30%. The discrimination index of the items were in the range of 0.90 – (-0.2) having a mean 0.40 with standard deviation of 0.33. 63% items were showing excellent discrimination index >0.35 followed by 27% with a poor i.e. <0.20 of D.I with some items also showing negative difficult index. Some studies have shown negative DI in 20% and 4% MCQ items.[11,15] Probable explanation was wrong key, ambiguous framing of questions or generalized poor preparation of students.[11] Items with negative DI decrease the validity of the test and should be removed from the collection of questions. The distractor efficiency of the items was in the range of 0-100 having a mean 59.21 with standard deviation of 31.45. 26.6% MCQ's were having 100% DE with 0 NFD. 7% MCQ's were having 0% D.E with 3 NFD. While almost 66.66% MCQ's were in between the range of 33.3% to 66.6% of D.E with 1 to 2 NFD's. Gajjar et al have shown that, in a total of 150 distracters, 133 (89.6%) were functional distracters, and 17 (11.4%) were NFDs. Items with NFDs were 15 (30%) out of which 13 items had DE of 66.6% and 2items had DE of 33.33%.[11] Students' performance depends on how distracters are designed.[16] The maximum discrimination (D = 0.5–0.6) was observed in acceptable range of difficulty index. i.e Diff -I (P = 30–70%). The discrimination index showed poor negative correlation with the difficulty index, which was found to be non-significant. (Pearson co-relation) r = -0.225, p-value=0.232. Difficulty index and

discrimination index are often reciprocally related except for extreme situations where the difficulty index is either too high or too low. It has been seen that the relationship between them is not linear, but predicted as dome shaped.[12,13] The findings of this study corroborated the same with maximum DI of items between p value of 40-60%. The overall reliability of our test was 0.61 suggesting it to be a less reliable test with poor internal consistency which is less than a study done by Mukherjee et al.[17] Though much data are not available regarding the reliability of the tests from various studies done on item and test analysis, one rule of thumb states that values greater than or equal to 0.70 are acceptable.[7,10]

CONCLUSION

As 80% items were having average difficulty and 63% having high and 27% having poor discrimination index, while 30% items having 2 to 3 non-functional distractor. It was concluded that item having average difficulty and high discriminating power with functional distracters should be incorporated into subsequent tests to serve the purpose it is meant to be.

Recommendations

Based on the findings and outcome of the above study we suggest conducting more and more studies in order to develop a valid question bank and to also identify the students scoring less mark. The students should be counselled personally by the staff to identify their difficulties and hence these problems must be dealt either by the modification of the teaching skills or by solving the difficulties the students are facing.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Kehoe J. Basic item analysis for multiple choice tests. PARE. 1995;4:10.
2. Zubairi AM, Kassim NLA. Classical and Rasch analysis of dichotomously scored reading comprehension test items. Malaysian J ELT Res. 2006;2:1-20.
3. Srivastava A, Dhar A, Agarwal CS. Why MCQ? Indian J Surg. 2004;66:246-8.
4. Carneson J, Delpierre G, Masters K. Designing and managing MCQs. Appendix C: MCQs and Blooms taxonomy, 2011. Available at: http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html. Accessed on 3 March 2017.
5. Case SM, Swanson DB. Extended matching items: a practical alternative to free response questions. Tech Learn Med. 1993;5:107-15.
6. Cizek GJ, O'Day DM. Further investigations of nonfunctioning options in multiple- choice test items. Educ Psychol Med. 1994;54(4):861-72.
7. Saudi Commission for Health Specialties. Item writing manual for multiple-choice questions. Available at: http: //www.scfhs.org.sa /education /HighEduExams/Guidlines/mcq/Documents/MCQ%20Manual.pdf. Accessed on 5 March 2017.
8. Mehta G, Mokhasi V. Item analysis of multiple choice questions – An assessment of the assessment tool. Int J Health Sci Res. 2014;4:197-202.
9. Shyam K, Pandey R. MCQ item analysis and Banking. The Art of Teaching Medical Students. 3rd Edition. New Delhi: Elsevier; 2015: 304-321.
10. Kuder–Richardson Formula 20. Available at: https://en.wikipedi.org/wiki/Kuder%E2%80%93Richardson_Formula_20 Accessed on 5 March 2017.
11. 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 J Comm Med. 2014;39:17-20.
12. Pande SS, Pande SR, Parate VR, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in Physiology. South East Asian J Med Education. 2013;7:45-50.
13. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices relationship between them in four response type multiple choice questions of pharmacology summative tests of year II MBBS students. Ie JSME. 2013;6:41-6.
14. Patel KA, Mahajan NR. Itemized Analysis of Questions of Multiple Choice Question (MCQ) Exam. Int J Sci Res. 2013;2:279-80.
15. Hingorjo MR, Laleel F. Analysis of One -Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. J Pak Med Assoc. 2012;62:142-7.
16. Dufresne RJ, Leonard WJ, Gerace WJ. Making sense of students answers to multiple choice questions. Phys Teach. 2002;40:174-80.
17. Mukherjee P, Lahiri SK, Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR J Dental Med Sci. 2015;14(12):47-52.

Cite this article as: Ingale AS, Giri PA, Doibale MK. Study on item and test analysis of multiple choice questions amongst undergraduate medical students. Int J Community Med Public Health 2017;4:1562-5.