Batna Journal of Medical Sciences

Item Analysis of Multiple Choice Questions used as a tool for Summative Assessment in a Medical College in Haryana

Analyse des éléments des questions à choix multiples utilisées comme évaluation sommative dans un institut médical à Haryana

Rashmi Gulia[1], Dr. Tapan Kumar Saha[2]

[1]1C-57, NIT Faridabad, Haryana – 121001, India

[2]Al-Falah School of Medical Sciences & Research centre Faridabad, Haryana- 121004

Corresponding author:
Rashmi GULIA
biochem.rg@gmail.com

DOI https://doi.org/10.48087/BJMSoa.2022.9105

Received : March 16, 2022
Accepted : May 13, 2022
Published : June 05, 2022

ABSTRACT

Introduction: Multiple choice questions (MCQs) are one of the popular tools used for evaluating medical students since they can be used to test a large number of candidates with only minimal human intervention. These MCQs also need testing for quality and standard which will ultimately improve the teaching-learning process. This study is conducted to assess the quality of MCQs of our in-use question bank. **Methods**: A total of 80 MCQs from the biochemistry internal examinations conducted in 2021 held at Al-Falah Medical College, Haryana for 147 students of 1st year Bachelor of Medicine and Bachelor of Surgery (M.B.B.S) were analyzed for difficulty index, discrimination index, and distractor effectiveness. The MCQs were the single best response type with one key option and three distractors. Statistical analysis for calculation of Mean, Standard Deviation, and Pearson correlation was performed on the data obtained using Microsoft excel 2010. **Results**: The observed mean of difficulty index, discrimination index, and distractor effectiveness was 75.08±15.68, 0.32±0.17, and 53.30±29.37 respectively. No item had a negative discrimination index. The Pearson correlation between DIF I and DI was negative (r = -0.43) and insignificant. **Conclusion**: Item analysis helps to optimize the existing question bank by helping in identifying flawed items. It is recommended to use this tool to standardize the testing process so as to accurately assess the student performance and improve the teaching-learning outcome.

Keywords: Medical education, Multiple-choice questions (MCQs), Item analysis, Difficulty Index, Discrimination Index

ABSTRACT

Introduction : Les questions à choix multiples (QCM) sont des outils populaires utilisés pour évaluer les étudiants en médecine, car peuvant être utilisées pour tester un grand nombre de candidats avec une intervention humaine minimale. Ces QCM doivent également être testés pour la qualité et la norme, ce qui améliorera en fin de compte le processus d'enseignement et d'apprentissage. Le but de cette étude est d'évaluer la qualité des QCM de notre banque de questions en cours d'utilisation. **Méthodes** : Un total de 80 QCM des examens internes de biochimie organisés en 2021 au Al-Falah Medical College, Haryana chez 147 étudiants de 1ère année de licence en médecine et de licence en chirurgie (M.B.B.S) ont été analysés afin évaluer l'indice de difficulté, l'indice de discrimination et l'efficacité du distracteur. Les QCM étaient le meilleur type de réponse unique avec une option clé et trois distracteurs. L'analyse statistique pour le calcul de la moyenne, de l'écart type et de la corrélation de Pearson a été effectuée sur les données obtenues à l'aide de Microsoft Excel 2010. **Résultats** : la moyenne observée de l'indice de difficulté, de l'indice de discrimination et de l'efficacité du distracteur était respectivement de 75,08 ± 15,68, 0,32 ± 0,17 et 53,30. ±29,37. Aucun item n'avait un indice de discrimination négatif. La corrélation de Pearson entre DIF I et DI était négative (r = -0,43) et non significative. **Conclusion** : L'analyse des éléments aide à optimiser la banque de questions existantes en aidant à identifier les éléments défectueux. Il est recommandé d'utiliser cet outil pour normaliser le processus de test afin d'évaluer avec précision les performances des élèves et d'améliorer les résultats d'enseignement et d'apprentissage.

Mots-clés : Formation médicale, Questions à choix multiples (QCM), Analyse d'items, Indice de difficulté, Indice de discrimination

INTRODUCTION

The teaching-learning process involves the teacher, the learner, and the curriculum, organized in systematic way to achieve the educational goals and objectives. Learning is the process of acquiring information whereas performance is the observed behavior /response. Assessment is a way of measuring learner's performance with the purpose to enhance student learning. It can be formative (detecting specific strengths and weaknesses of individual, guiding future learning, providing reassurance, promoting reflection, and shaping values) or summative (making an overall judgement about competence, fitness to practice, or qualification for advancement to higher level of responsibility) [1].

An ideal tool for evaluation which can satisfy all the requirements has not been devised yet. Assessment differs on the level it being done (cognitive, psychomotor or affective domain), the need for doing so, the type (formative or summative), the form (criterion referenced or norm-referenced assessment) or the skill being assessed [1, 2]. The objective type questions offer a distinct advantage of being more reliable as compared to essay type questions – not only because of predetermined correct

Citation :

Gulia R, Saha TK. Item analysis of single- best response type Multiple Choice Questions used as a tool for summative assessment. Batna J Med Sci 2022;9(1):24-27. https://doi.org/10.48087/BJMSoa.2022.9105

answer but also because more content can be tested in shorter duration. They can be used repeatedly and are easy to mark. Objective questions are marked objectively but they are prone to subjective errors. Therefore, due care and attention has to be given while formulating and writing these questions or items. Each item should test an important learning outcome. The question part of the MCQ is called 'stem'; the correct answer is called the 'key' and the remaining questions are called as 'distractors' [3-7]. The quality of an MCQ depends upon the distractors- i.e. how often they are able to distract the student who does not know the correct answer. Thus, a thoughtful and deliberate attempt has to be made to provide effective distractors [2].

A good item has DIF I of 50%, which means approximately half the class can answer it correctly [2]. Questions/Items that are too easy or too difficult should not be included in a test unless there are sufficient strong reasons to do so. Items with high DI are useful for the purpose of selection of students as in an entrance exam. In the situations where ranking is required the items with good positive discrimination and moderate difficulty are chosen whereas to assess classroom performance, the items with high difficulty and low discrimination indices are useful to ensure that all students have learnt the lesson correctly [1].

Summative assessment conducted at the end of unit/ term/semester/course helps to determine the achievement of a satisfactory level of competence on the part of the learner. It is used for making pass/fail decisions in most set-ups and is also known as assessment of learning [2]. Multiple choice questions (MCQs) are one of the popular tools used for evaluating medical students, since they can be used to test large number of candidates with only minimal human intervention. These MCQs also need testing for quality and standard. Item analysis is a method of testing each individual test items (MCQ) for the quality, which is beneficial for both students and faculty [3]. Multiple choice questions if constructed properly have the ability to assess higher cognitive processing like interpretation, analysis, and problem solving of Bloom's taxonomy instead of just recall of facts. [4]. A single best response (SBR) type of MCQ is the one in which the student is asked to select the single correct/best answer from a list of choices. Each MCQ consists of a stem (the problem) and the responses (key option and distractors) [1, 2]. The aim of the study is to assess the quality of MCQs used for end of term internal examination and optimize the test items of our in-use question bank.

MATERIALS AND METHODS

This study was conducted as a part of assessment in the Department of Biochemistry of at Al-Falah School of Medical Science & Research Centre, Faridabad, Haryana. Students of 1st year MBBS had appeared for two internal examinations conducted in October and November 2021. Each exam comprised of two question papers with each paper having 20 MCQs making up a total of 80 "single best response type" MCQs. All MCQs had one stem with four responses of which one was key option and other three were distractors. One mark was awarded for each correct response but there was no negative marking for incorrect response. We had made four sets of paper by disorganizing the sequence of questions so as to avoid the possibility of copying among neighboring students.

Data analysis: Item analysis was done for post-validation of the questions in the paper. Data obtained was entered in MS excel 2010 and analyzed. The scores of 147 students was arranged in order of merit and was then divided into three groups. The upper one-third were considered as high achievers (H) group and lower one-third as low achievers (L) group [1, 2]. Out of these 147 students, 49 were in H group and 49 were in L group. Scores of these two groups were used for item analysis. The remaining 49 in the middle group were set aside and not considered for analysis. Total 80 MCQs and 240 distractors were analyzed based on this data. Each item was analyzed for difficulty index (DIF I), discrimination index (DI) and distractor effectiveness (DE) using standardized formulae [1, 5].

DIF I is the proportion of students in the two groups who have answered the item correctly. It is calculated by DIF I = [(H+L)/N] x 100, where H is the number of students answering the item correctly in high achievers group and L is the number of students answering the item correctly in low achievers group and N is the total number of students in the two groups including non-responders. DIF I ranges between 0 - 100%. Items with higher value of DIF I (>70%) are considered too easy and that with lower DIF I (<30%) are too difficult. Difficulty level of 30-70% is considered acceptable and usually 50-70% is recommended [1, 2].

DI is a measure of ability of the item to discriminate between knowledgeable and ill-informed students and is calculated by DI = 2 x [(H-L)/N]. Item discrimination describes how effectively the test item differentiates between high ability and low ability students, with higher DI values being desirable. The value of DI ranges between 0 and 1, where DI of 1 is ideal. Items with DI >0.35 are considered excellent. The minimum acceptable DI is taken as 0.25 for practical purposes where one of the four SBR type MCQs are used. DI of 0.25-0.35 is good and those with DI between 0.15-0.25 needs to be revised. Clearly, the items with DI < 0.15 are defective. A negative DI (DI < 0) indicates a wrong item/key and is undesirable [1]. It happens when the students with lower ability select the correct response by guess work without having actual knowledge and in turn outdo the good students [1, 2, 5]

DE measure the functionality of alternatives to the correct response in the item, as the lack of functionality would lead to a greater possibility of guessing the correct answer. Distractors chosen by < 5% of students were considered as nonfunctional distractor (NFD) [1, 2, 6]. DE was calculated based on the number of NFDs per item. It ranges from 0-100%. If an item contains three, two, one or zero NFDs, then DE will be 0, 33.3%, 66.6% and 100% respectively [5].

Items were categorized according to their values obtained for various indices - DIF I, DI and DE. The data are reported as percentage and mean ± standard deviation. Further, the corrective actions were proposed based on the values obtained. The relationship between DIF I and DI was measured using Pearson's correlation test. A value of $P < 0.05$ was considered to be statistically significant.

RESULTS

A total of 147 students appeared for the internal examination consisting of 80 MCQs. The mean DIF I, mean DI and mean DE was 75.08±15.68, 0.32±0.17, and 53.30±29.37 respectively as depicted in Table 1.

The DIF of 59 (73.75%) MCQs were easy (DIF I >70%), 2 (2.5%) was difficult (DIF I <30%) and the remaining 19 (23.75%) items were in acceptable range (DIF I 30-70%). The DI for these items showed 12 (15%) items with poor discrimi-

Table 1. Assessment of difficulty index, discrimination index and distractor effectiveness for 80 items amongst 147 students.

Parameter	Mean	Standard Deviation
Difficulty Index (DIF I)	75.08	15.68
Discrimination Index (DI)	0.32	0.17
Distractor effectiveness (DE)	53.30	29.37

nation ability (DI <0.15), and 34 (42.5%) items having excellent discrimination (DI > 0.35). There were 16 (20%) items with good discrimination index (DI 0.25-0.35) and 18 (22.5%) were in the category where there is a need to be revise them (DI 0.15-0.25). There were 240 distractors for 80 MCQs. Amongst these, 13 (16.25%) MCQs had functional distractors and rest 67 (83.75%) MCQs had either one, two or three non-functional distractors (NFDs). The distribution between DIF I and DI is depicted in Table 2. And, the DE analysis is shown in Table 3. The Pearson correlation between DIF I and DI was negative (r= -0.43) and non-significant with a p-value > 0.05, this relationship can also be observed from the scatter plot shown in Figure 1.

Table 2. Distribution of items between difficulty index and discrimination index.

Parameter	N = 80	Interpretation	Action
Difficulty Index (DIF I)			
< 30	2 (2.5%)	Difficult	Revise/Discard
30-70	19 (23.75%)	Acceptable	Store
> 70	59 (73.75%)	Easy	Revise/Discard
Discrimination Index (DI)			
< 0.15	12 (15%)	Poor	Discard
0.15-0.25	18 (22.5%)	Revise	Revise
0.25-0.35	16 (20%)	Good	Store
> 0.35	34 (42.5%)	Excellent	Store

DISCUSSION

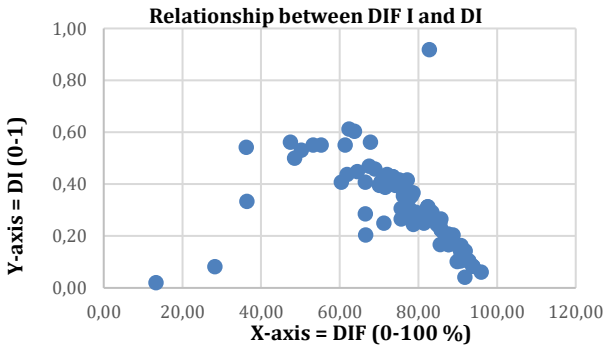
Each item/MCQ being used for assessment must be evaluated so that the assessment process does not remains flawed. Post-test examination of the MCQs helps to assess the quality of each test item and the overall test as a whole [6]. This exercise provides feedback to the educators on the potential actions to be taken for improvement of existing set of questions. In our study, the mean DIF I was 75.08 ± 15.68 %, with 73.75 % of MCQs falling in the too easy category of the difficulty index (DIF I >70%), 23.75 % had acceptable difficulty (DIF I 30-70 %) and only 2.5 % were in the too difficult category (DIF I < 30%). Rao et al. had mean DIF I as 50.16 ± 16.15 with 85% items in acceptable range, Kumar et. al had mean of 55.32 ± 7.4 with 82 % items in acceptable range. These observations are in line with study by Patel and Mahajan with 80% items in acceptable range [3, 7-8]. Karelia et. al have shown the range for mean DIF I between 47.17 ± 19.79 and 58.8 ± 19.33 in a study conducted by them over a period of 5 years with 61% items in the acceptable range [9]. Some of the previous studies have shown mean difficulty index as 63.06 ± 18.95, 39.4 ± 21.4, 52.53 ± 20.59, 48.90 ± 13.72, 59.18 ± 15.14 and 57.92 ± 19.58 [6, 10-14]. Higher the difficulty index lower is the difficulty of the question. Too difficult items can lead to deflated scores, while too easy items may result into the inflated scores and a decline in motivation level among students. So, the items with high DIF I (>70%) can be placed at the start of the test as "warm-up" questions to boost the confidence of students or may be discarded altogether. Similarly items with low DIF I (<30%) should be either revised or removed completely. [1, 13]. Our study had only 2 items which were too difficult and 59 were too easy and rest 19 items were within the acceptable range of difficulty index.

Table 3. Analysis of functional and non-functional distractors.

Distractor Analysis	(N=240)
Functional distractors (FDs)	39 (16.25%)
Non-functional distractors (NFDs)	201 (83.75%)
Items with 3 NFDs (DE=0%)	9 (11.25%)
Items with 2 NFDs (DE=33.3%)	31 (38.75%)
Items with 1 NFDs (DE=66.6%)	27 (33.75%)
Items with 0 NFDs (DE=100%)	13 (16.25%)
Mean ± Standard deviation of DE	53.30±29.37

DI of an item indicates its ability to differentiate between knowledgeable and ill-informed student. The value of discrimination index normally ranges between 0 and 1 [1, 4]. Our study had mean DI of 0.32 ± 0.17, with 62.5 % items having acceptable to excellent DI. These findings were in line with study by Rao et al., with mean DI of 0.34 ± 0.17 and 85 % items having acceptable to excellent level of DI [3]. Other studies by Bhat et al., Gajjar et al., Mehta et al., Kumar et al., and Kaur et al. have shown mean DI of 0.26 ± 0.15, 0.14 ± 0.19, 0.33 ± 0.18, 0.31 ± 0.12, and 0.37 ± 0.15 respectively [4-7,13]. Karelia et al., have shown a mean DI ranging from 0.29-0.38 with 78% items having an acceptable to excellent DI of ≥ 0.2 for a study on 200 test items of 10 MCQs analyzed from the year 2008-2012 [9]. In another study by Singh et al. on item analysis of 20 MCQs, they have similarly reported 70% items with DI ≥ 0.20 [15]. The mean DI of study by Kheyami et. al on items analyzed from 2013 to 2016 was 0.20-0.34 [16]. A study by Patel and Mahajan on item analysis of 50 items, have reported 82 % items having DI ≥ 0.20 [8]. A negative DI can also occur in case both high and low achiever groups answer the MCQ by guesswork. There were no item/MCQ having negative discrimination index in current study. Items with negative discrimination index should be removed as they decrease the validity of a test. The DIF I and DI are reciprocally related. In our study, the Pearson correlation between DIF I and DI was small and negative (r= -0.43) and not significant at P<0.05 similar to Bhat et al., with a small negative correlation of -0.27 [4]. Karelia et al., reported insignificantly (P < 0.05) small positive correlation (r=0.11) along with two other studies [9-11].

Figure 1. Scatter plot showing relationship between Difficulty index and Discrimination index with a negative Pearson's correlation of r= -0.43 and p value >0.05.



In our study of 240 distractors, there were a total of 201 (83.75%) non-functional distractors (NFDs) and 39 (16.25 %) were functional distractors. Amongst these, 11.25 % items had poor DE (DE=0%), 38.75 % had moderate DE (DE = 33.3%), 32.5 33.75 % had good DE (DE= 66.6%) and remaining 16.25 % had excellent DE (DE =100%). Similar other studies

by Rao et al., Gajjar et. al., Bhat et al., and Kumar et al. have reported 5%, 11.3%, 21.6% and 27% non-functional distractors in their respective studies [3-5, 7]. Patil et al., Kaur et al., and Kolte et al. have also reported 14.06%, 18%, 7.5% NFDs respectively in their studies [12-14]. The current study has reported a comparatively higher percentage of non-functional distractors. These were analyzed and corrective action were taken subsequently to improve the quality. Conducting similar studies time to time would be helpful to strengthen the MCQ bank.

CONCLUSION

Item analysis is done with the aim to standardize the testing process so that a student does not get either more or less marks than he actually deserves. The non-standardized testing is ultimately going to harm the learning – by eroding the faith of students in the system and also by producing incompetent doctors per se. DIF I, DI and DE are commonly employed for the analysis of an item with the purpose of detecting specific technical flaws in the question and to get information for improving them. The item analysis of internal examination conducted by the department of Biochemistry indicated that majority of test items had acceptable to easy level of difficulty index. Although these items had good discrimination index, but most of these had non-functional distractors. An item analysis is simple and worthy procedure which is recommended in order to continuously improve the quality of the question bank. Since, the National Medical Commission (Erstwhile Medical Council of India) has made it mandatory to include MCQs in the formative and summative assessment of the medical graduates, as per the competency based medical curriculum (CBME), it is essential that the testing process be standardized to accurately assess student performance. Special training / workshops should be offered to the faculty members in order to develop skills for preparing effective MCQs and the ability to optimize them as and when required.

CONFLICT OF INTEREST: There are no conflicts of interest.

ACKNOWLEDGEMENT: Mr. Nizammudin has helped with the compilation of the study data.

REFERENCES

1. Saha TK, Datta K, editors. Essentials of medical teaching technology. New Delhi: Army College of Medical Sciences; 2010.
2. Tejinder S, Piyush G, Daljit S. Principles of Medical Education. 4th ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd; 2013.
3. Rao C, Prasad HK, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. International Journal of Educational and Psychological Researches. 2016 Oct 1;2(4):201. DOI: 10.4103/2395-2296.189670 https://www.ijeprjournal.org/text.asp?2016/2/4/201/189670
4. Bhat SK, Prasad KH. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. Indian Journal of Ophthalmology. 2021 Feb;69(2):343. Indian J Ophthalmol. 2021 Feb;69(2):343-346. doi: 10.4103/ijo.IJO_1610_20.PMID: 33463588
5. 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 Ahmedabad, Gujarat. Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine. 2014 Jan;39(1):17. PMID: 24696535 PMCID: PMC3968575 DOI: 10.4103/0970-0218.126347
6. Mehta G, Mokhasi V. Item analysis of multiple choice questions-an assessment of the assessment tool. Int J Health Sci Res. 2014 Jul;4(7):197-202. DOI: 10.18502/sjms.v16i3.9695 ISBN: 1858-8530
7. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. Item analysis of multiple choice questions: a quality assurance test for an assessment tool. Medical Journal Armed Forces India. 2021 Feb 1;77:S85-9. doi: 10.1016/j.mjafi.2020.11.007. Epub 2021 Feb 2.PMID: 33612937
8. Patel KA, Mahajan NR. Itemized analysis of questions of multiple choice question (MCQ) exam. Int J Sci Res 2013;2:279-80. DOI: 10.15373/22778179/FEB2013/94
9. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of year II M.B.B.S students. IeJSME 2013;7:41-6. IeJSME 2013 7(2): 41-46
10. Ananthakrishna N. The item analysis: In: Medical education principles and practice; by Anathakrishnan N, Sethukumaran KR, Kumar S, editors. ch. 20. Pondicherry, India: Alumni Association of National Teacher Training centre, JIPMER. 2000:131-7.
11. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in physiology. South-East Asian Journal of Medical Education. 2013 Jun 30;7(1):45-50. DOI: 10.4038/seajme.v7i1.149 ISBN: 2536-905
12. Patil VC, Patil HV. Item analysis of medicine multiple choice questions (MCQs) for under graduate (3rd year MBBS) students. RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES. 2015 May 1;6(3):1242-51. ISBN: 0975-8585
13. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. International Journal of Applied and Basic Medical Research. 2016 Jul;6(3):170. PMID: 27563581 PMCID: PMC4979297 DOI: 10.4103/2229-516X.186965
14. Kolte V. Item analysis of multiple choice questions in physiology examination. Indian J of Basic & Applied Medical Research. 2015;4(4):320-6. P ISSN: 2250-284X , E ISSN : 2250-2858
15. Singh JP, Kariwal P, Gupta SB, Shrotriya VP. Improving multiple choice questions (MCQs) through item analysis: An assessment of the assessment tool. Int J Sci Appl Res. 2012;1:53-7.
16. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. Sultan Qaboos University Medical Journal. 2018 Feb;18(1):e68. DOI: 10.18295/squmj.2018.18.01.011 ISBN: 2075-051X