

Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study

Subrahmanya K Bhat, Kishan HL Prasad

Purpose: Multiple-choice questions (MCQs) are useful in assessing student performance, covering a wide range of topics in an objective way. Its reliability and validity depend upon how well it is constructed. Defective Item detected by item analysis must be looked for item writing flaws and optimized. The aim of this study was to evaluate the MCQs for difficulty levels, discriminating power with functional distractors by item analysis, analyze poor items for writing flaws, and optimize. **Methods:** This was a prospective cross-sectional study involving 120 MBBS students writing formative assessment in Ophthalmology. It comprised 40 single response MCQs as a part of 3-h paper for 20 marks. Items were categorized according to their difficulty index, discrimination index, and distractor efficiency with simple proportions, mean, standard deviation, and correlation. The defective items were analyzed for proper construction and optimized. **Results:** The mean score of the study group was 13.525 ± 2.617 . Mean difficulty index, discrimination index, and distractor efficiency were 53.22, 0.26, and 78.32, respectively. Among 40 MCQs, twenty-five MCQs did not have non-functioning distractor; 7 had one, 5 had two, and 3 had three. Of the 20 defective items, 17 were optimized and added to the question bank, two were added without modification, and one was dropped. **Conclusion:** Item analysis is a valuable tool in detecting poor MCQs, and optimizing them is a critical step. The defective items identified should be optimized and not dropped so that the content area covered by the defective item is not kept of the assessment.

Key words: Difficulty index, discrimination index, item analysis, nonfunctioning distractor, optimization

The medical education across the world consists of the initial assessment of learner's need, monitoring the teaching-learning activities, certification of the competence to award a degree and practise medicine in context to the need of the society.^[1] Student learning, certification and quality assurance are the aims of assessment.^[2] It also acts as a strong incentive for learning.^[1-3] Properly constructed multiple-choice questions (MCQs) can assess higher cognitive processing like interpretation, analysis, and problem solving of Bloom's taxonomy instead of just recall of facts.^[3-5] It was found to be superior to the modified essay question in assessing higher-order skills.^[6] Keeping in view the widespread use of MCQs in the assessment of medical students, this study was undertaken for Item analysis and to identify the difficulty and discrimination indices and distractor efficiency. Defective MCQs were analyzed and optimized based on item writing guidelines.^[1,7]

The aims of this study were to evaluate MCQs for difficulty levels and discriminating power with functional distractors by item analysis and to analyze the items with poor indices for Item writing flaws and optimize them for a viable question bank in Ophthalmology.

K S Hegde Medical Academy, Deralakatte, Mangalore, Karnataka, India

Correspondence to: Dr. Subrahmanya K Bhat, K S Hegde Medical Academy, Deralakatte, Post Nithyananda Nagar, Mangalore - 575 018, Karnataka, India. E-mail: drsbhatk@yahoo.co.in

Received: 24-May-2020

Revision: 23-Jul-2020

Accepted: 01-Aug-2020

Published: 18-Jan-2021

Access this article online

Website:

www.ijo.in

DOI:

10.4103/ijo.IJO_1610_20

Quick Response Code:



Methods

It was a prospective, cross-sectional study involving MBBS students taking Ophthalmology examination. Forty MCQs were part of the final formative assessment, which included a 3-h Question paper for 100 marks with long essays, short essays, short answer questions, and MCQs. After taking permission from the Dean of the college and consent from study participants, 120 students were assessed in Ophthalmology. The MCQs for the same were chosen randomly and were prevalidated by two departmental colleagues. The 40 MCQ items of 'single response type' having a stem, a key and three distractors were chosen. Each correct response (key) was awarded "half mark," an incorrect response or un-attempted Item was awarded "Zero mark." Time allotted for attempting the MCQs was initial 30 minutes. Strict invigilation was carried out to avoid malpractices. The faculties have evaluated the MCQs with the correct keys. After scoring, students were ranked in order of merit using Microsoft Excel 2010. Bottom third (40) of low achievers (L) and top third (40) of high achievers (H) were used for item analysis to serve the twin purpose of having groups large enough to be representative and

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

Cite this article as: Bhat SK, Prasad KH. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. Indian J Ophthalmol 2021;69:343-6.

different enough to be meaningful.^[1] Each Item was analyzed for difficulty index (DIF I), discrimination index (DI), and distractor efficiency (DE) using the standard formulae.^[1] DIF I is the percentage of students who answered the Item correctly and given by $DIF I = [(H + L)/N] \times 100$. DI is a measure of the effectiveness of an item in discriminating between L from H and is given by $DI = 2 \times [(H-L)/N]$, where H is the number of students answering the Item correctly in the high achieving group, L is the number of students answering the Item correctly in the low achieving group, and N is the total number of students in the two groups (including non-responders). Distractors which were chosen by less than 5% of students were considered as a nonfunctional distractor (NFD).^[1] DE was determined for each Item based on the number of NFDs in it and ranged from zero to 100%. DE of an item was considered as 100% if there were no NFD and 66.6%, 33.3% and 0% if it contained 1, 2, or 3 NFDs, respectively.^[8] The relationship between the DIF I and DI were determined by Pearson correlation coefficient (r) using an online calculator (<https://www.socscistatistics.com/tests/pearson/>). P value of <0.05 was considered to be statistically significant.

Items were categorized according to their DIF I, DI, and DE with simple proportions, mean, standard deviations, and correlation. Data were interpreted, as shown in Table 1.^[1] The range of values for DIF I is 0%–100%. The range of values for DI is –1.0 to +1.0. The DI becomes negative if more students in the L group answer correctly than those in the H group.

Results

The score obtained by the study group ranged from 6.5 to 20, with a mean of 13.53 ± 2.62 . Mean of DIF I, DI, and DE was 53.22 ± 22.44 , 0.26 ± 0.16 and 78.32 ± 32.11 , respectively. The distribution of DIF I and DI of the items and their corresponding DE with actions proposed are shown in Tables 2 and 3, respectively. The scattered diagram [Fig. 1] shows the relationship between the DIF I and DI. DI was poor for both easy and difficult items. Pearson correlation between DI and DIF I was small and negative ($r = -0.207$) and not significant at $P < 0.05$ ($P = 0.199$). The relationship of NFDs with mean DIF I and DI of items are shown in Table 4.

Item analysis showed 20 items (50%) to be defective, and the subject expert safely added the remaining 20 items to the question bank with an item card for each Item. Of 20 defective items, 17 items were found to have bad stem or distractors,

and these were optimized and added to the question bank. It included five easy items with normal DI, five easy items with poor DI, one easy Item with negative DI, two difficult items with normal DI, one difficult item poor DI, two items with acceptable difficulty but poor DI, 1 item with good DI and DIF I but two defective NFD. One difficult Item and one Item with acceptable difficulty having poor DI did not have any flaws and added to the question bank without alteration. One difficult Item with negative DI was dropped because there was one more question, based on the same data in the same paper with a hint to the correct answer. Feedback was given to the faculty on all difficult items for corrective measures in teaching and framing the Item. All optimized items were appropriately tagged via the item card for future reassessment.

Discussion

The Item analysis should be regularly carried out for creating a good question bank updating recent advances so that MCQs can be used in the evaluation of the cognitive skills of medical students effectively.^[9-16] The MCQs will provide feedback to the teachers on their educational actions. Designing MCQ is a complicated and time-consuming process in a multidisciplinary and integrated curriculum. In this study, the mean DIF, DI, and DE were 53.22 ± 22.43 , 0.26 ± 0.15 , 78.31 ± 32.11 , respectively, which were in the acceptable range, and a few other studies reported similar results.^[10-12,15] In this study, 60% of items were of acceptable difficulty (DIF I = 30-70%), 27.5% items easy, and 12.5% difficult. Similar results were seen in a few of the studies.^[9,12,14] Nearly 80% of the items were in an acceptable range in three studies.^[10,11,17] In studies by Kheyami D (53.4%),^[15] Shenoy PJ (40%),^[18] Gajjar S (48%),^[8] and Rajkumar P (46.6%),^[13] it was less. In the last

Table 1: The DIF I and DI interpretation

DIF I	Interpretation	DI	Interpretation
>70%	Easy	>0.35	Excellent
30%-70%	Good	0.2-0.35	Good
<30%	Difficult	<0.2	Poor

Table 2: Distribution of items according to DIF I with DE and actions proposed (n=40)

DIF I	Interpretation	No. of items (%)	DE	Action taken
≥ 70	Too easy	11 (27.5)	39.37	Optimized and stored
30-70	Good	24 (60)	91.65	1 item optimized and stored, others stored unaltered
<30	Too difficult	5 (12.5)	100	1 discarded, 3 optimized and stored, 1 stored unaltered

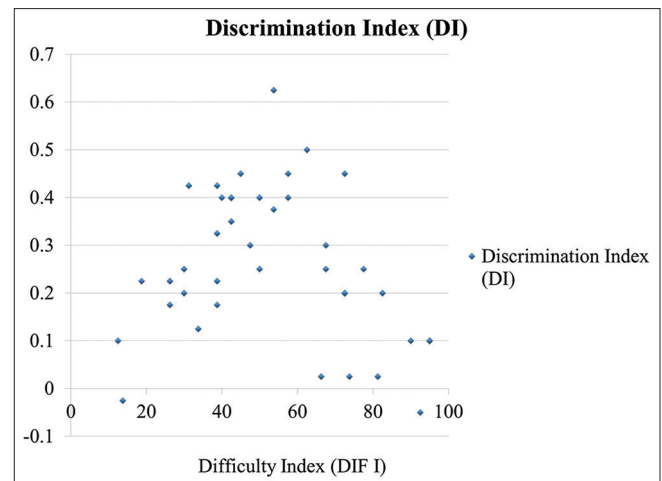


Figure 1: Relationship between Difficulty Index and Discrimination Index

Table 3: Distribution of items according to DI with DE and action taken (n=40)

DI	Interpretation	No. of items (%)	DE	Action
<0.2	Poor	10 (25)	63.32	1 discarded, 2 stored unaltered, others optimized and stored
0.2-0.35	Good	15 (37.5)	84.42	Stored unaltered
DI >0.35	Excellent	13 (32.5)	87.17	1 optimized and stored, others stored unaltered

Table 4: Relationship of Non-functioning distractors (NFDs) with DIF I and DI (n=40)

Parameter	Items with 0 NFDs	Items with 1 NFD	Items with 2 NFDs	Items with 3 NFDs
Number (%)	25 (62.5%)	7 (17.5%)	5 (12.5%)	3 (7.5%)
Mean DIF I (%)	30.4	58.03	74.25	94.17
Mean DI	0.287	0.254	0.245	0.05

two studies, the acceptable range was defined as 30%–60%, unlike our and other (30%–70%) studies.

In this study, 28 (70%) items had good DI (>0.2). Similar findings were seen in a few studies^[14-16] but was less compared to studies by Karelia BN (78%),^[9] Kaur M (86%),^[17] Shenoy PJ (80%),^[18] Pande SS (75%),^[12] Rao C (85%)^[10] and Mozaffer RH (88%).^[11] In our study, 15 (37.5%) items had one or more NFD which was similar to a study by Kaur M (38%)^[17] but less compared to study by Mozaffer RH (58%),^[11] Rajkumar P (43.3%)^[13] and Shenoy PJ (65%).^[18] In our study, of 120 distractors, 26 (21.6%) were non-functional, which is comparable to previous studies.^[11,13,17] More NFD was seen by Shenoy PJ (33.3%),^[18] but less NFD was observed by Gajjar S (11.3%)^[8] and Rao C (5%).^[10] Seven items with two or more NFD and four items with 1 NFD with poor DI/DIF I was optimized and added the question bank. Three items with 1 NFD had normal DI, DIF I, and distractors and were added to the question bank without alteration. One of Item with 2 NFD, had good DI and DIF I but distractors were defective and were optimized and added to the question bank. One of the easy Item with multiple writing flaws with a negative DI and 3 NFDs re-establishes the finding by Omer AA that flawed items affected high achievers more than low achievers.^[19]

The defective Items identified by item analysis should be appropriately analyzed for flaws, and it should be optimized. These should not be blindly dropped; otherwise, some skills may be left out of the assessment.^[9-11] In early formative assessments with fewer topics to be studied and short term memory assessed, majority of students may answer the Item correctly (DIF I >70%) reducing the DI complimenting an efficient teaching-learning process. Decreased or negative DI can also occur with poor teaching-learning interaction with DIF I <30%. In this scenario, both high and low achievers answer the MCQs by guesswork. DI decreases when the Item becomes too easy or too difficult. This was noted in this study where maximum discrimination (0.4) was seen when the DIF I is 40-60%, and the relationship was dome-shaped as found in many studies.^[9,12,14-16,20] In our study, Pearson correlation between DI and DIF I was small and negative ($r = -0.207$) and not significant at $P < 0.05$ ($P = 0.199$). A study by Mitra *et al.* also showed a moderate negative correlation in their study ($r = -0.325$, significant at $P < 0.01$).^[16] Karelia *et al.*,

showed a small positive correlation ($r = 0.11$, insignificant at $P < 0.05$).^[9] Two other studies showed a slight positive correlation significant at $P < 0.01$.^[12,15] In this study, items that were not acceptable due to poor DIF I, DI, or bad distractors were analyzed for writing flaws and optimized keeping in mind the guidelines.^[1,7]

In this study, some of the Item writing flaws were unique to Ophthalmology. The presence of two conflicting terms in options which cannot co-exist, like mydriasis and miosis, lid retraction and ptosis, intumescent lens, and shrunken lens suggesting that one of them has to be chosen or excluded even to a low achiever was one of the writing flaws. Other flaws detected were *All of the above* as the key, the *EXCEPT* in the stem not capitalized and made bold, options used were not uniform, presence of abbreviations in the stem and options, the part of the word same in stem and options such as phacomorphic and phacoanaphylaxis. Few others being, questions related to controversial information in the books recommended for undergraduates and distractors which are unrelated and implausible. One of the easy Item (DIF I = 92.5%) with negative DI and 3 NFD which low achievers answered better was Bitot spots are seen in a) Vitamin A deficiency b) Dry eye c) Sjogren syndrome d) Lagophthalmos. It was optimized as 'Bitot spot can be seen in all the following *EXCEPT* a. Cirrhosis of the liver, b. Celiac disease, c. Dietary deficiency of vitamin A, d. Primary Sjogren syndrome' which tests the same cognitive domain in a better way. Defective Items need to be appropriately reconstructed; validated and feedback should be given to the faculties for corrective action. The New Graduate medical regulation 2019 released by the Medical Council of India has made it mandatory to include MCQs in the formative and summative assessment in a competency-based medical curriculum.^[21] We hope item analysis will serve as a helpful tool to generate question banks at departmental and university levels which will provide items with acceptable difficulty and discrimination indices.^[5-10] Our study has included only 40 MCQs. Periodical Item analysis with more MCQs is necessary for validation of viable question bank in important subjects like Ophthalmology.

Conclusion

To conclude, Item analysis is a valuable tool in detecting poor MCQs, and optimizing them is a critical step. It will help us to identify poorly constructed items and give attention to optimize them to improve the quality of the question bank. With nearly half of the items detected to be less than optimum, efforts to optimize the stem and distractor is essential, without which the purpose of the assessment will be defeated. It also points toward the need to have more frequent faculty development programs in creating standard MCQs, effective pre-validation, and to have viable question bank in ophthalmology.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Ananthakrishnan N. Multiple choice questions facts and fantasies and Item Analysis validation and banking of MCQs. In: N Ananthakrishnan, K R Sethuraman, Santhosh Kumar, editors. *Medical Education Principles and Practice*. 2nd ed. Pondicherry: Alumni Association of National Teacher Training Centre JIPMER; 2000. pp. 119-37.
- Mutch A, Brown G. Assessment: A guide for heads of department. In Chapter 2, What is an assessment strategy? LTSN Generic Centre, York: Learning and teaching supporting network; 2001. pp. 4-6. Available from: https://www.academia.edu/33257467/Assessment_A_guide_for_Heads_of_Department. [Last accessed on 2020 May 17].
- Carneson J, Delpierre G, Masters K. Designing, and managing multiple choice questions. In Chapter 2, What are Multiple Choice Questions? 2nd ed. University of Cape Town; 2016. pp. 3-6. Available from: https://www.researchgate.net/publication/309263856_Designing_and_Managing_Multiple_Choice_Questions_2nd_Ed?channel=doi&linkId=58074fef08ae03256b783474&showFulltext=true. [Last accessed on 2020 May 17].
- Al Muhaissen SA, Ratka A, Akour A, AlKhatib HS. Quantitative analysis of single best answer multiple-choice questions in pharmaceuticals. *Curr Pharm Teach Learn* 2009;11:251-7.
- Khan MUZ, Aljarallah BM. Evaluation of modified essay questions (MEQ) and Multiple choice questions (MCQ) as a tool for assessing the cognitive skills of undergraduate medical students. *Int J Health Sci* 2011;5:39-43.
- Palmer EJ, Devitt PG. Assessment of higher-order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. *BMC Med Educ* 2007;7:49.
- Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ* 2002;15:309-34.
- 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 J Community Med* 2014;39:17-20.
- Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and the relationship between them in four-response type multiple choice questions of pharmacology summative tests of Year II MBBS students. *IeJSME* 2013;7:41-6.
- Rao C, Kishan Prasad HL, Sajitha K, Permi H, Shetty J. Item analysis of multiple-choice questions: Assessing an assessment tool in medical students. *Int J Educ Psychol Res* 2016;2:201-4.
- Mozaffer RH, Farhan J. Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *JPMA* 2012;62:142-7.
- Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in the formative exam in physiology. *SEAJME* 2013;7:142-7.
- Rajkumar P, Sachin BP, Kamesh V, Abhijit Vinod B. Evaluation of multiple-choice questions by item analysis in a medical college at Pondicherry, India. *Int J Community Med Public Health* 2016;3:1612-6.
- 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.
- 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 Univ Med J* 2018;18:68-74.
- 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;3:2-7.
- Kaur M, Singla S, Mahajan R. Item analysis of in-use multiple choice questions in pharmacology. *Int J App Basic Med Res* 2016;6:170-3.
- Shenoy PJ, Sayeli V, Rao RR. Item-analysis of multiple choice questions: A pilot attempt to analyze formative assessment in pharmacology. *Res J Pharm Biol Chem Sci* 2016;7:1683-7.
- Omer AA, Abdulrahim ME, Albalawi I. Flawed multiple-choice questions put on the scale: What is their impact on students' achievement in a final undergraduate surgical examination?. *J Health Spec* 2016;4:270-5.
- 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;35:67-71.
- Regulations on Graduate medical education (Amendment), 2019. Board of governors in the suppression of the Medical Council of India. New Delhi. 2019. pp. 1-99.