Item Analysis of Multiple-Choice Questions in Pharmacology in an Indian Medical School

Manoor Narasimha Sachidananda Adiga[1], Swathi Acharya[1], Rajendra Holla[1]

[1]Department of Pharmacology, K.S. Hegde Medical Academy, Deralakatte, Nitte (Deemed to be) University, Mangaluru, Karnataka, India

Address for correspondence Swathi Acharya, MBBS, MD, Department of Pharmacology, K.S. Hegde Medical Academy, Deralakattte, Nitte (Deemed to be) University, Mangaluru 575018, Karnataka, India (e-mail: swathiacharya@nitte.edu.in).

Abstract

Introduction   Student assessment by multiple-choice questions (MCQs) is an integral part of student evaluation in medicine. The medical teacher should be trained to construct an item with proper stem and valid options. Periodic item analyses will make the process of assessment more meaningful. Hence, we conducted the study to analyze MCQs (item analysis) tested on a batch of MBBS students in pharmacology in their three internal assessment examinations.

Methods   The study was conducted in the Department of Pharmacology of a medical college in Mangaluru on 150 students. The MCQs of the three internal assessment examinations (20 each) respectively were analyzed. We analyzed each question for difficulty index (DI), discrimination index (DsI), and distracter efficacy or functionality and expressed the percentage results.

Results   The DI was in an acceptable range of 60, 75, and 90%, respectively, in the three internal assessments. The percentage of "too difficult" questions was 10, 20, and 10% and the average DsI was 0.32 ± 0.04, 0.28 ± 0.02, and 0.26 ± 0.02, respectively. In the second and third internal assessments, 95% of questions had functional distracters, while in the first internal assessment, only 60% of questions had functional distracters.

Conclusion   We conclude from our study that even though the items (MCQs) framed for the internal assessments were in the acceptable range of quality in terms of the parameters assessed, we must improve MCQ's construction in selecting distracters in some topics.

Keywords
► item analysis
► MCQs
► difficulty index
► discrimination index

Introduction

Many methods can assess the learning and competency of undergraduate students. In the professional courses, it is essential to evaluate the competence acquired during their professional training. The understanding of the subject is an integral part of performing and mastering the competency. Hence, it is crucial to assess the student's understanding of the subject in professional courses. Therefore, objectivizing the evaluation has become increasingly important in education for summative and formative purposes. One such method is multiple-choice questions (MCQs), which is being very commonly employed nowadays.[1]

The MCQ-based evaluation assesses the knowledge, evaluates understanding, and analyzes student's power.[2] An

MCQ consists of a stem, a complete or incomplete statement, followed by four to five options with a single best answer. Constructing a right stem with appropriate options needs experience and sound knowledge on the subject. These well-constructed items will be preferred over other methods since they can assess the students objectively with minimal bias from the assessor and the comprehensive coverage of the subject.

Item analysis is the process of analyzing an MCQ's performance after it has appeared in an examination. The item analysis evaluates the question on three parameters. The difficulty of the questions that were asked can be analyzed by judging the difficulty index (DI) or facility value of the item. The discrimination index (DsI) measures the ability of the item to discriminate good students from others. The item analysis is also useful to get feedback on the functionality of the alternative or distracter efficacy (DE), which gives the idea of the quality of the distracters compared with the correct response.[1] "Item analysis" examines student responses to individual test items (MCQs) and test as a whole.[3] It helps to improve/revise items and the test.[4] Ideally, the constructed item can assess the cognitive, affective, as well as the psychomotor domains.[2,4-7]

Periodic assessment of item analysis on different batches will enable the teachers to have a pool of "good question banks," which are considered ideal, and get feedback on valid MCQ construction. It also guides the teachers to fine-tune their method of teaching wherever needed. The present study's objective is to analyze the MCQs that were tested in internal assessment examinations of pharmacology in a batch of MBBS students to assess them critically and take remedial corrective measures if required in future objective assessments.

Materials and Methods

In this retrospective study conducted in a medical college in Mangaluru, Karnataka, we considered MCQs tested in the three internal assessment pharmacology examinations (20 each) in a batch of (150) MBBS students for item analysis. Each assessment session contained 20 MCQs (item), with each item carrying 4 options with a single best response type (key). The students were given 20 minutes to answer these questions, and there was no negative marking for the wrong answer. The master chart was prepared in which students' performances with their responses to each item were entered. The students were ranked based on their MCQ score, and they were divided into three equal groups. The top one-third was assigned as high achievers, and bottom one-third as low achievers.

We analyzed each item for DI, DsI, and DE or functionality of the distracters[1] with the following formula:

$$P = [(H + L)/N] \times 100 \text{ (expressed as percentage)},$$

where

P = Facility value or DI.

H = Number of students answering the item correctly in the high-achieving group.

L = Number of students answering correctly in the low-achieving group.

N = Total number of students in the two groups, including the nonresponders.

The DI (or P) of each MCQ in the three assessments was expressed as follows:

- When P was >70%, MCQ was considered "easy."
- When P was 30–70%, MCQ was considered of "acceptable difficulty level."
- When P was <30%, MCQ was considered as "too difficult."

The DsI[1] was calculated by the formula:

$$DsI = 2 \times [(H-L)/N]$$

A value greater than 0.35 was considered "excellent," between 0.20 and 0.34 was considered "good," and less than 0.20 was considered as "poor."

Distracter Efficacy:[1] This is the number of distracters in each item across the entire batch that has not attracted even 5% of responses. This is expressed as percentage of functional distracters, that is, 0, 33.3, 66.6, and 100% for an item with 3, 2, 1, and 0 distracters not selected by <5% students, respectively.

Statistical Analysis

The average DI, DsI, and the DE were expressed as percentage mean ± standard error of the mean for each assessment. The different categories in each parameter were also expressed in percentages.

Ethical Considerations

The study was performed after obtaining the institutional ethics committee approval.

Results

Among the 150 students, 108 students responded in all the 3 sessions included for the analysis.

Difficulty Index: The percentage of easy questions in different internal assessments was 30, 5, and 0%, respectively. The percentage of too difficult questions was 10, 20, and 10% in three internals, respectively. The percentage of items at an acceptable level of difficulty was 60, 75, and 90%, respectively, in the three assessments (►**Fig. 1**). The average DI of items as a whole was 60.69 ± 4.53, 43.54 ± 3.54, and 38.45 ± 2.21% in the three internal assessments (►**Table 1**), which was in the acceptable test range (30–70%).

Discrimination Index: The average DsI in the three internal assessments was 0.32 ± 0.04, 0.28 ± 0.02, and 0.20 ± 0.02, respectively (►**Table 1**). Overall, 55% of questions in the first internal were categorized as excellent, 20% in the second, and 30% in the third internal assessment. Also, 25% of questions were considered acceptable in the first internal, 60% in the second, and 50% in the third internal. In the first internal, 20% of questions were categorized as poor, 30% in the second, and 20% in the third, as
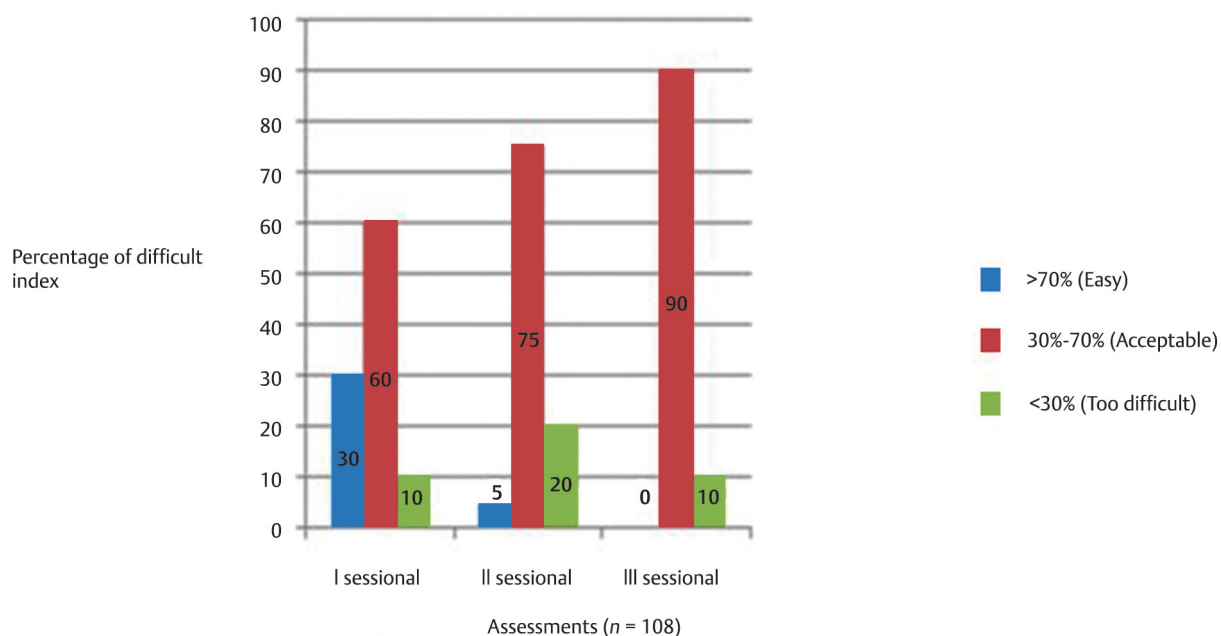
Fig. 1 Percentage of difficulty index in all three internal examinations.

Table 1 Consolidated results of item analysis of multiple-choice questions in three internal assessments

S. no.	First internal assessment			Second internal assessment			Third internal assessment		
	DI (%)	DsI	DE (%)	DI (%)	DsI	DE (%)	DI (%)	DsI	DE (%)
Q1	29.48	0.37	100	41.82	0.29	100	34.55	0.25	100
Q2	87.96	0.16	33.3	11.82	0.16	100	52.73	0.33	100
Q3	32.41	0.43	100	72.73	0.33	100	33.64	0.38	100
Q4	64.81	0.37	100	36.36	0.25	100	28.18	0.24	100
Q5	77.78	0.11	66.6	30.00	0.31	100	31.82	0.24	100
Q6	36.11	0.43	100	46.36	0.34	100	31.81	0.38	100
Q7	28.74	0.54	100	55.45	0.31	100	38.18	0.36	100
Q8	51.85	0.37	66.6	21.82	0.29	100	54.55	0.36	100
Q9	92.59	0.07	0	52.73	0.33	100	34.55	0.36	100
Q10	76.85	0.24	66.6	59.09	0.27	100	40.91	0.24	100
Q11	55.55	0.41	100	45.45	0.25	100	35.45	0.27	100
Q12	47.22	0.61	100	60.00	0.14	100	45.45	0.22	100
Q13	98.15	0	0	45.45	0.36	100	17.27	0.20	100
Q14	58.33	0.31	100	22.73	0.13	100	57.27	0.27	66.66
Q15	53.70	0.52	100	48.18	0.49	100	30.00	0.05	100
Q16	48.15	0.41	100	20.91	0.02	100	39.09	0.05	100
Q17	65.74	0.46	66.6	46.36	0.27	100	40.91	0.38	100
Q18	74.07	0.30	66.6	65.45	0.4	66.66	38.18	0.18	100
Q19	69.44	0.31	100	41.82	0.22	100	51.82	0.31	100
Q20	62.96	0.26	100	46.36	0.42	100	32.73	0.14	100
Mean ± SD	60.69 ± 4.53	0.32 ± 0.04	78.31 ± 7.74	43.54 ± 3.54	0.28 ± 0.02	98.33 ± 7.37	38.45 ± 2.21	0.26 ± 0.02	98.33 ± 7.37

Abbreviations: DE, distracter efficacy; DI, difficulty index; DsI, discrimination index; SD, standard deviation.
Notes: n = 108. Questions in all three assessments were different.

they failed to discriminate between the good and mediocre students (►**Fig. 2**).

Functionality of Distracter (or DE): In the first internal assessment, eight questions had options not selected by even 5% of students. One option in five questions was not chosen in these eight questions by less than 5% of students. Two options in one question were not recognized as answer by less than 5% of students. In two questions, three options were not chosen by less than 5% of students (►**Table 1**, ►**Fig. 3**). In the case of second and third internals, only one question had an option that was not chosen by the student as the answer (►**Table 1**; ►**Fig. 3**).

Discussion

MCQs are an integral part of assessing medical students objectively. Properly constructed questions serve to determine the subject understanding of the students. Time-to-time analysis of these items will help retain the good ones and discard the improper ones. Item analysis will also help teachers construct good questions. An item is usually analyzed for the difficulty level of the question, DsI, and functionality of each distracter chosen for a question.

The DI will provide the overall picture of the MCQ test. Item's key is considered acceptable if the DI ranges between 30 and 70%, which means 30 to 70% of the upper-third and lower-third students in the class correctly answered that MCQ. If the DI is more than 70% for an MCQ, it is considered easy (contradictory), and when it is less than 30%, that MCQ is deemed to be difficult. Ideally, MCQs should be in the acceptable range.

In our study, the average DI of the test varied from 38.45% in the third internal assessment to 60.69% in the first internal assessment. Pande et al reported that the mean DI was 39.4% and 52.5% in their study.[8] As per Karelia et al, it was 47.17% and 58.08% in the analysis of two assessments.[9] Hence the DI of our examinations was at par with the other similar item analyses. In our study, the acceptable range items were 50, 75, and 90%, respectively, in the three internals. This was similar to findings in other studies: 69 in Karelia et al,[9] 80 as per Patel and Mahajan,[10] and 62% as per Mehta and Mokhasi.[11] The rest of the items, that is, 50, 25, and 10%, were unacceptable (either too easy or too difficult). These results were again in line with other study findings: 31,[9] 20,[10] and 38%.[11]

The DsI is a measure of the item to discriminate between students with higher and lower abilities and ranges between 0 and 1. It identifies good students from mediocre students. In general, the value between 0.20 and 0.35 is considered as good. Items with DI > 0.35 are considered excellent, while those with DI < 0.20 are considered poor. The DsI is inversely related to the DI because, if the DI value is higher, the item poorly discriminates the students with good abilities from poor abilities. Conversely, the lower DI value differentiates two categories of students.[11] The DsI sometimes is a negative value; it means that the students with lower abilities have guessed the answer correctly. Students with higher abilities had some doubt about that item or some flaws in teaching that particular concept. Items of such kind should be addressed explicitly in future classes by emphasizing or employing suitable methods such as small group teaching.[3,4]

In our study, the average DsI was 0.32 ± 0.04, 0.28 ± 0.02, and 0.26 ± 0.02 in the three assessments, respectively (►**Table 1**). Items with good to excellent DI were 80% each in the three assessments in our study (►**Fig. 2**), which was similar to the results described by Mehta and Mokhasi in which the DsI was 0.33 ± 0.18.[11] The finding was in contrast to Gajjar
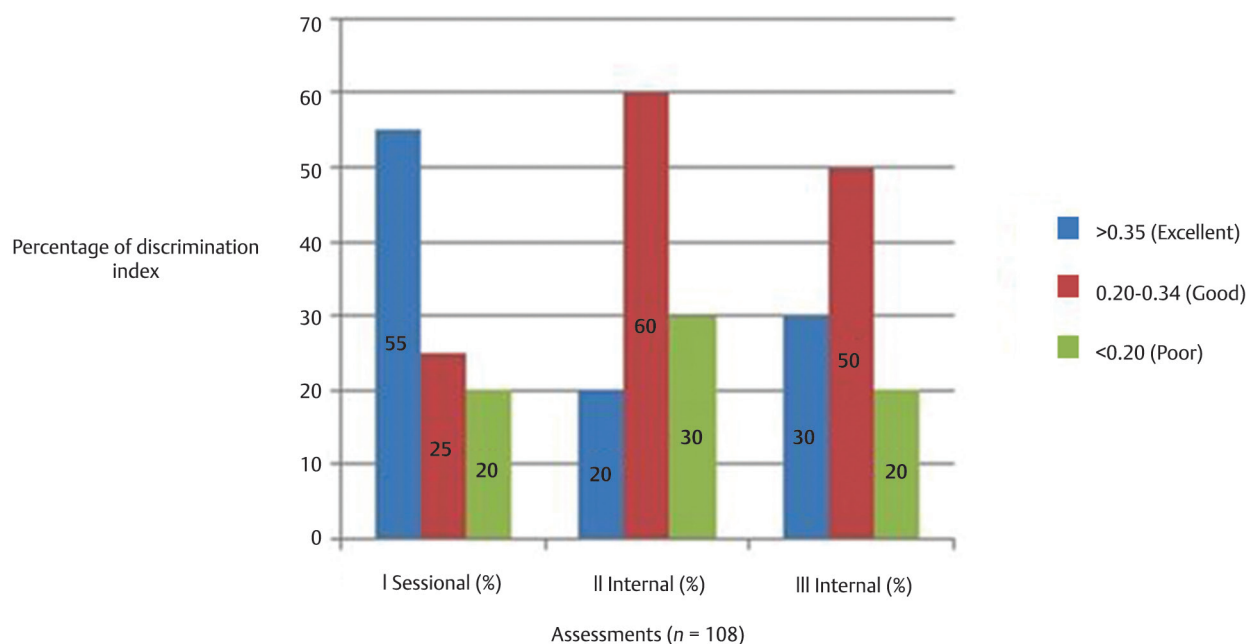

Fig. 2 Percentage of discrimination index in all three internal examinations.
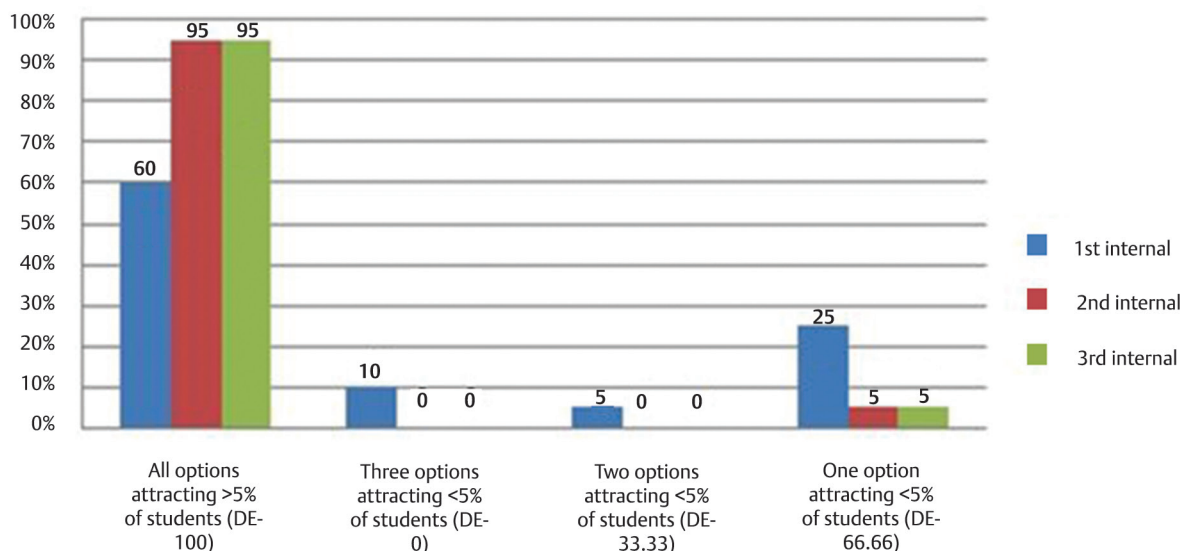
Fig. 3 Distracter efficacy in percentage in all three internal examinations.

et al's findings, where DI was 0.14 ± 0.19, and 48% of items were considered good to excellent.[12]

The functionality of distracter, or DE, is a measure of each option's ability to distract the student, or, in other words, each distracter is perceived by the students as an answer so that all the options are chosen by at least 5% of students as the answer. Most of the items will have a key (correct response) and three options, which are distracters. Functional or useful distracter refers to the option selected by more than 5% of students. In comparison, the nonfunctional distracter is the option other than the key, chosen by less than 5% of students. DE for an item ranges from 0 to 100. If an item has three non-functional distracters, then the DE of that item is 0, while DE is 33.33, 66.66, and 100% if it contains 2, 1, and nil nonfunctional distracter.[3,4,6]

In our study, the average DE of the three assessments was 78.31 ± 7.74, 98.33 ± 7.37, and 98.33 ± 7.37, respectively. In total, out of 60 items, 50 items (>83%) had 100% DE, 7 items had 66.6%, and 1 item had 33.33%, while 2 items had 0% DE. The conclusion drawn is that items constructed in our assessments were excellent, except for two items across the three assessments. It was mainly in the first internal where 25, 5, and 10% of items had DE of 66.66, 33.33, and 0%, respectively. In the second and third internal assessments, only one item had a DE of 66.66%. In comparison, the rest of the items had 100%. It indicated the need for precaution to be taken while constructing distracters, which can increase the DE of the items. Gajjar et al reported that 70% of items had 100% DE in their study, while the rest of the items had DE ranging from 33.33 to 66.66%, with no items having 0% DE.[12] Different studies have quoted varying degrees of functional distracter ranging from 18.66,[12] 52.2,[6] and 89.6%.[13]

Ideal objective assessment by MCQ requires careful construction of items that reflect on the evaluation of students with different abilities; hence student's performance can be taken as a yardstick of the quality of the assessment. Analysis of the items in terms of various parameters mentioned above identifies the error in constructing an item that may be revised, replaced, or removed if deemed.

Limitations of the Study: Being a retrospective study, the number of items analyzed for each internal assessment was small. The items analyzed in our study were from the internal assessments, which were not compared with the level of different parameters for items of university examination questions.

Conclusion

We conclude from our study that all the assessments had an acceptable range of DI, and good DsI. The DE, with few exceptions, was also good. Therefore, constructing ideal MCQs with good distracters requires more effort from the teachers. Serially designed item analysis exercise will significantly help in having a good pool of good-to-excellent items in terms of DsI and DE.

Conflict of Interest
None declared.

Acknowledgments

References

1 Ananthakrishnan N, Medical Education – Principles and Practice. In: Ananthakrishnan N, Sethuraman KR, Kumar S, eds. Item Analysis: Validation and Banking of MCQs. 2nd ed. All India press; 2000:131–137

2 Scorepak®: Item analysis. www.washington.edu./oea/score1/htm. Accessed April 13, 2013

3 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 ; pp. 70–77

4 Matlock-Hetzel S. Basic concept in item and test analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, January, 1997. www.ericae.net/ft/tamu/espy.html. Accessed April 13, 2013

5 Sarin YK, Khurana M, Natu MV, Thomas AG, Singh T. Item analysis of published MCQs. Indian Pediatr 1998;35(11):1103–1105

6 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;9:Article 40 doi: 10.1186/1472-6920-9-40

7 Scantron Guides - Item Analysis, adapted from Michigan State University website and Barbara gross devils tools for teaching. www.freepdfdb.com/pdf/item-analysis-scantron. Accessed April 13, 2013

8 Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in formative exam in physiology. South East Asian J Med Educ. 2013;7:45–50

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 MBBS students. IeJSME. 2013;6:41–46

10 Patel KA, Mahajan NR. Itemized analysis of questions of multiple choice question (MCQ) exam. Int J Sci Res (Ahmedabad) 2013;2:279–280

11 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

12 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(1):17–20

13 Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc 2012;62(2):142–147