

ANALYSIS OF THE PERFORMANCE OF MCQs AS A PART OF FORMATIVE ASSESSMENT FOR 1ST MBBS STUDENTS IN BIOCHEMISTRY

Suryakar Vrushali Prabhunath¹, Surekha T. Nemade², Ganesh D. Ghuge³

¹Associate Professor, Dept. of Biochemistry, Hind Institute of Medical Sciences, Safedabad, Barabanki, Uttar Pradesh – 225003.

²Associate Professor, Dept. of Biochemistry, Dr. Vasant Rao Pawar Medical College, Hospital and Research Centre, Nashik, Maharashtra – 422003.

³Professor & Head, Dept. of Biochemistry, Dr. Vasant Rao Pawar Medical College, Hospital and Research Centre, Nashik, Maharashtra - 422003.

Article Info: Received 28 August 2019; Accepted 14 September. 2019

DOI: <https://doi.org/10.32553/ijmbs.v3i9.534>

Address for Correspondence: Dr. Suryakar Vrushali P., Associate Professor, Dept. of Biochemistry, Hind Institute of Medical Sciences, Safedabad, Barabanki, Uttar Pradesh – 225003.

Conflict of interest: No conflict of interest.

Abstract

Introduction: Multiple Choice Questions (MCQs) is one of the most preferred tool of assessment in medical education as a part of formative as well as summative assessment. MCQ performance as an assessment tool can be statistically analysed by Item analysis. Thus, aim of this study is to assess the quality of MCQs by item analysis and identify the valid test items to be included in the question bank for further use.

Materials and methods: Formative assessment of 1st MBBS students was carried out with 40 MCQs as a part of internal examination in Biochemistry. Item analysis was done by calculating Difficulty index (*P*), Discrimination index (*d*) and number of Non-functional distractors.

Results: Difficulty index (*P*) of 65% (26) items was well within acceptable range, 7.5% (3) items were too difficult whereas 27.5% (11) items were in the category of too easy. Discrimination Index (*d*) of 70% (28) items fell in recommended category whereas 10% (4) items were with acceptable, and 20% (8) were with poor Discrimination index. Out of 120 distractors 88.33% (106) were functional distractors and 11.66% (14) were non-functional distractors. After considering difficulty index, discrimination index and distractor effectiveness, 42.5% (17) items were found ideal to be included in the question bank.

Conclusion: Item analysis remains an essential tool to be practiced regularly to improve the quality of the assessment methods as well as a tool for obtaining feedback for the instructors.

Key Words: Difficulty index, Discrimination index, Item analysis, Multiple choice questions, Non-functional distractors

INTRODUCTION

Assessment is a critical component of instructions, especially in the field of medical education where it is generally said that “Assessment drives the learning”¹ and thus playing a central role in determining what and how students learn. If properly used, it can aid in accomplishing key curricular goals. Tests are a strong motivator and are a way to convey students what is important. Assessment also helps to fill instructional gaps by providing the feedback to the teachers as well as encouraging students to read broadly on their own.² Tests have a powerful influence on students’ future learning even as health professionals expanding their professional horizons. And thus, it is important to develop tests or assessments that will

further educational goals.³ In view of this it becomes important to review or analyse the process of assessment itself.

Multiple Choice Questions (MCQs) are commonly used as a part of formative as well as summative assessments for undergraduate students since last two decades in our evaluation settings of MBBS curriculum. MCQs has its’ own pros and cons when compared to other types of question formats. MCQs are versatile, allows assessment of different cognitive domains like Knowledge, Comprehension, Application, Analysis and are reliable as a form of assessment,⁴ provided they are structured properly which itself is a challenging and time consuming process.⁵

For any assessment method to be effective it has to be valid, reliable, properly constructed and in alignment with its purpose and most importantly it has to be devoid of any flaws.⁶ We can improve the quality of assessment method by eliminating the flaws and improving the quality of items itself and analysing the items' actual performance. Item analysis helps us to grade or review the quality of MCQ based on its' difficulty index i.e. the difficulty level, discrimination index i.e. ability to discriminate a good student from not so good student and distractor effectiveness.^{7, 8} Haladyana in a study showed that more than 50% test items constructed by educators failed to produce desired results.⁹ Thus item analysis is essential, so that flawed or ineffective items can be identified and either revised or detained from further use in order to improve the quality of assessment test.

Aim and Objectives:

Thus, this study was undertaken with the aim to evaluate the performance of assessment by item analysis and to bank the "ideal" items for future use. This is achieved by following objectives:

1. To perform the item analysis by calculating
 - a) Difficulty Index – i.e. whether the item is of appropriate level of difficulty for the batch of students tested at that level.
 - b) Discrimination Index – i.e. whether the item is capable of discriminating between the knowledgeable and the ill-informed students.
 - c) Distractor efficiency/effectiveness
2. To find the correlation between Difficulty Index and Discrimination Index
3. To identify the items of known difficulty and discrimination indices which can be used for banking.

Materials and Methods

The study was conducted in the Department of Biochemistry, at Dr. Vasantrao Pawar Medical College, Adgaon, Nashik, Maharashtra. 40 single best response type of MCQs administered for the submission test as a formative assessment in Biochemistry to the 1st MBBS students were analysed. The MCQ items were constructed by all the teachers in the Biochemistry department and were scrutinized at the department for content accuracy as a usual practice. Out of all items, 40 scrutinized questions were selected by the Head of the Department and

formatted for question paper as a formative assessment after completing the respective interactive didactic sessions. Each item consisted of a stem with four options, one correct response (key) and other three alternatives (distractors). Students were informed at least 3 weeks prior regarding the syllabus, date, time and venue of the assessment test.

Out of 100 1st MBBS students, 96 students appeared for the assessment test. We scored the whole test for all the students. A correct response was awarded 0.5 marks and the wrong one zero marks, no negative marks were awarded. Students were ranked in order of merit based on their test scores.

Top third (H) (n=32) were taken as high achievers and bottom third (L) (n=32) as low achievers. These two groups were considered for calculating difficulty and discrimination index of each item. The data was entered in Microsoft excel sheet and indices were calculated using Microsoft office excel 2017. Each item was analysed for its difficulty index, discrimination index and distractor efficiency with the help of following formulae:^{7,8,10-15}

1. **Difficulty Index (P):** The item difficulty or difficulty index or facility value indicated by the symbol 'P' is calculated by the formula

$$P = \frac{H + L}{T} \times 100$$

T

Where,

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

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

T = Total number of students in two groups including non-responders.

P value or difficulty index is interpreted as follows:

P < 30% – Difficult

P = 30% to 70% - Acceptable

P > 70% - Easy

2. **Discrimination Index (DI or d):** It is the measure of the ability of the item to discriminate between good/actually knowledgeable students and not so good/ill-informed students.

The discrimination index 'd' is calculated by the formula

$$d = \frac{H - L}{T} \times 2$$

Where H, L & T are same as above.

Discrimination Index (d) is interpreted as follows:

- d = negative – defective item/wrong key
- d < 0.15 – Poor discriminating ability
- d = 0.15 to 0.25 – Acceptable
- d = 0.25 to 0.35 – Good discriminating ability
- d > 0.35 – Excellent discriminating ability

Distractor Efficiency (DE): Non-functional distractor (NFD) in an item is the option or alternative (other than the key), selected by <5% students whereas the alternative selected by >5% students is called as Functional distractor (FD).¹⁶ Depending on the number of NFDs present in an item, its’ distractor effectiveness (DE) ranges from 0% to 100%. If an item contains three or two or one or nil non-functional distractor (NFD) then DE will be 0%, 33.3%, 66.6%, and 100%, respectively.⁷ We analysed 120 distractors for distractor efficiency.

Pearson’s Correlation “r” was calculated for Difficulty index and Discrimination index.

Observations and Results

Total 40 items of single best response type MCQs and 160 options (40 correct answers i. e, key and 120 distractors) were analyzed. The results are illustrated in the tables.

Table 1: Distribution of 40 items as per their difficulty index (facility value ‘P’) with its’ interpretation:

Difficulty Index/ Facility Value “P” Value (in %)	< 30	30 To 70	> 70
No. of MCQs	3	26	11
Percentage of Items	7.5%	65%	27.5%
Interpretation	Difficult (or Defective)	Acceptable	Easy

Mean = 55.94% S.D. = 18.39%

Table 2: Distribution of 40 items as per their discrimination index (DI or d) with its’ interpretation:

Discrimination Index (“d” value)	<0.15	0.15 – 0.25	0.25 – 0.35	>0.35
No. of MCQs	8	4	13	15
Percentage of Items	20%	10%	32.5%	37.5%
Interpretation	Poor	Acceptable	Good	Excellent

Mean = 0.30 S.D. = 0.17

For an item to be acceptable it should have difficulty value (P) in the range of 30 to 70 and a discrimination index (DI or d) more than 0.25, thus accordingly when the two indices were combined 25 items i.e. 62.5% of the total items were “Ideal”.

Distractor Analysis:

For distractor analysis, out of 120 distractors for 40 items, 106 (88.33%) distractors were functional distractors whereas 14 (11.66%) distractors were found to be non-functional distractors (NFDs). 29 items had 100% distractor efficiency i.e. nil NFDs and 9 items had single non-functional distractor (DE=66.66%), whereas one item had two NFDs (DE= 33.33%) and one was found to have all the three distractors as NFDs (DE= 0%)

Table 3: Distribution of 40 items as per their Distractor Efficiency (D.E.):

No. of NFDs	D.E.	No. of Items	Percentage (%) of Items
0	100%	29	72.5
1	66.66%	9	22.5
2	33.33%	1	2.5
3	0%	1	2.5

Correlation between Difficulty Index and Discrimination Index:

Pearson Correlation between difficulty index and discrimination index showed no significant correlation (r=0.16).

Discussion

As stated by Boud D¹⁷ “Students can, with difficulty, escape from the effect of poor teaching. However, they cannot (if they want to graduate), escape the effects of poor assessment.” Assessment being one of the most fundamental, judgmental and influential phenomenon in the teaching and learning of the students, needs continuous evaluation.¹⁸

Item analysis helps to identify the quality of items which in turn decides the quality of assessment test, which is one of the key elements in curriculum development.¹⁹ Hence each item has to be evaluated for its’ difficulty index, discrimination index and distractor efficiency not only to determine quality of assessment but also for further use and banking of items.

In this study 40 items of single best response type MCQs with 120 distractors from the Biochemistry, administered for a formative assessment to 96 1st MBBS students were evaluated.

Difficulty Index (P): Difficulty Index (*P*) or Facility Value refers to the percentage of students getting the item correct and thus as mentioned earlier, items with difficulty Index (*P*) < 30 % will be more difficult and those with *P* > 70% will be easier. In our study the mean difficulty index was found to be 55.94±18.39% which is well within the acceptable range of 30-70% and is consistent with previous other similar studies where it ranges from 43 to 65%.^{11-13, 18, 20, 21} We found 65% (26) items were in the acceptable range (*P* between 30-70%), whereas 27.5% (11) items were too easy i.e, *P* > 70 and 7.5% (3) items were too difficult. (*P* < 30).

When the *P* is very small it is indicative of a difficult question. This could be due to improper coverage of the area related to test item or may be that part was difficult for the students to grasp. It also may indicate that the area assessed is not appropriate at that level of formative assessment for the students.²²⁻²⁴ Thus items with *P* <30% needs to examine for construction of the stem, presence of any confusing language, ambiguous distractors or may be an incorrect key and should be revised or discarded otherwise. But sometimes it is advisable to have some difficult questions also in test paper to select the toppers. Similarly easy questions can be retained to increase the confidence of students.

Discrimination Index (d): Our study demonstrated the mean DI (*d*) of 0.30 ± 0.17 which is comparable to other similar studies demonstrating the mean DI ranging from 0.21-0.37.^{11-13,18,20-21} We had 28 items with good/excellent DI (*d* > 0.25), out of these 15 items (*d* > 0.35) were excellent in discriminating the actual knowledgeable students versus ill-informed students.

In this study though 4 items (*d* = 0.15 to 0.25) needs to be revised, whereas 8 MCQs (*d* < 0.15) are to be discarded. Sometimes we can use items with low DI (*d*) for adequate or representative sampling of the course content and objectives. Thus the purpose of the assessment will affect the discriminating power of the particular item and hence the item discrimination indices are always interpreted in relation to the type of test being analysed.²⁵

As in the study carried out by Saxena et al¹² we also had 20% (8) items with *d* <0.15 with two items having negative DI. Poor discrimination index commonly indicates either ambiguous wordings or wrong stem or key or it may indicate grey areas of understanding.^{26,27} So when we further analysed these two items with negative discrimination index we found that one had a wrong key and other was a defective stem due to a typographical error, and both these items had *P* <30. Matlock-Hetzel²⁸ presented that items with negative DI has to be discarded as these are not only useless, but they also reduce the validity of the test.

Relationship between the 'P' and Difficulty Index (d):

Difficulty index and discrimination index are usually inversely related.²² But, we could not found any such relationship in our study. Rao et al¹¹ in a similar study found that moderately easy/difficult (acceptable range) items had the maximal discriminative ability. Items with *P* values closer to 0.50 are considered more useful in differentiating between individuals.²⁹

Distractor Analysis:

Distractor efficiency (*DE*), is one such tool that indicate whether the item was well constructed or failed to perform its purpose, ultimately affecting the quality and purpose of the assessment itself.¹⁶ In our study with 40 MCQs and 120 distractors we found 106 (88.33%) distractors were functional distractors (FDs) whereas 14 (11.67%) were non-functional distractors (NFDs). Mean *DE* was 88.33 ± 22.07 and is comparable with other studies.^{11,12,20,21} Saxena et al¹² found that, out of 90 distractors, 86 (95.55%) were functional distractors and only 04 (4.44%) were NFDs, and the mean *DE* was 95.55 ± 11.55 whereas Rao et al¹¹ demonstrated the distraction effect of items was 89.99%. Suryakumar Namdeo¹³ in a study from Bhubaneswar showed that out of 75 distractors, 40 (53.4%) were NFDs and 35 (46.4%) were FDs. In this study they observed 32% items with one NFD, 40% items had two NFDs whereas 16% items had 3 NFDs.

The number of NFDs also affect the discriminating power of an item. It has been observed that decreasing the number of distractors from four to three decreases the difficulty index while increases the discrimination index and reliability. In our study there were 29 items with nil NFDs, 9 items with one NFD and there was one item with 2 NFDs and one item with 3 NFDs. The items with 2 or 3 NFDs needs

to be revised or replaced, but the items with a single NFD may be used. Tarrant et al¹⁶ has concluded that items with two functioning distracters were more difficult than items with three functioning distracters. Whereas Hingorjo³⁰ observed that items having one NFD had excellent discrimination ability ($d = 0.427$) as compared to items with all four functioning distracters ($d = 0.351$). But again as we commonly use 3 distractors with one correct option, the fact that “items having one NFD has better results in items having 3 distractors”, needs to be evaluated.

Thus, functionality of the distracters in itself is an independent factor affecting the quality and item functioning. Hence, designing or constructing the distractors with equal plausibility is an important but at the same time is a difficult task. More the number of non-functional distractors easier the item, also increases the P , whereas more the number of functional distractors P value of the item decreases making the item more difficult.

After considering all the indices of item analysis (P , d and DE) of the given assessment, out of 40 single best response MCQs as items there were 17 (42.5%) items which can be considered as ideal and hence can be used for banking for future use. 7 items having single NFD may or may not be revised and rest 16 items needs to be revised or replaced.

Conclusion

The functionality of distracters, the flaws in item writing, and the optimum number of options are interrelated and affect the item quality, item performance, and the test performance as a whole. Item analysis remains as an essential tool to validate the reliability and feasibility of a MCQ based assessment. It helps in indicating the assessment quality and assessing the functionality of the test carried out. It helps in identifying the gaps in the instructions and further improvements in instructions, thus serving as a very effective feedback mechanism for teaching as well. Item analysis provides a measure to identify the flawed items and thus provides an opportunity to construct better quality items. It also helps in identifying the items with average difficulty, with good to excellent discrimination index having functional distractors which we call as Ideal items and are banked for future use.

Thus this study reinforces the fact that item analysis is an essential tool to be practiced regularly as a

routine in all the formative as well as if possible in summative assessments too, to improve the quality of assessment methods which could further the learning and curricular goals of medical students.

References

1. Wormald BW, Schoeman S, Somasunderman A. Assessment drives learning: an unavoidable truth? *Anat Sci Educ* Oct 2009; 2 (5):199-204.
2. Epstein R. Assessment in medical education. *N Engl J Med* 2007; 356:387-96. <http://dx.doi.org/10.1056/NEJMra054784>
3. Friedman BDM. The role of assessment in expanding professional horizons. *Med Teach* 2000; 22:472-7. <http://dx.doi.org/10.1080/0142159005011073>
4. Walsh CM, Seldomridge LA. Measuring Critical thinking. *Nurse Educ.* 2006 ;31(4):15942.
5. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. *Natl Med J India* 2008; 21(3):130-3
6. Vander Vleuten C P M, Scheuwirth L W T. Assessing professional competence: from methods to programmes. *Med Educ.* 2005; 39: 309-317.
7. 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.
8. Manual of Basic Course in Medical Education Technologies, Maharashtra University of Health Sciences, Regional Training Centre, Aundh, Pune. 2013; 45-51.
9. Haladyna TM. MC formats. In : Haladyna TM (ed.). *Developing and validating multiple-choice test items.* 3rd Edition. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2004. P 67-96.
10. Singh T, Gupta P, Singh D. Test and Item Analysis. In: *Principles of Medical Education.* Fourth edition. New Delhi Jaypee Brothers Medical Publishers (P) Ltd; 2013:108-13.
11. Rao C, Kishan Prasad H. L, 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.* Oct-Dec 2016; 2 (4): 201-04.
12. Saxena S, Srivastava PC, Mallick AK, Joshi HS, Das B. Item analysis: An unmatched tool for validating MCQs in Medical Education. *Indian Journal of Basic and Applied Medical Research.* Sep 2016; 5 (4): 263-269
13. Namdeo SK, Sahoo B. Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar India. *Int J Res Med Sci* May 2016; 4(5) :1716 – 1719.
14. Thorndike RM, Cunningham GK, Thorndike RL, Hagen EP. *Measurement and evaluation in psychology and*

- education. 5th edition. New York, NY, England: MacMillan Publishing Co, Inc; 1991
15. Hotiu, A. The relationship between item difficulty and discrimination indices in multiple-choice tests in a Physical science course (MSc thesis). Florida Atlantic University, Boca Raton, Florida; 2006. Available on: www.physics.fsu.edu/research/education/A.Hotiu_thesis.pdf
 16. 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: 40.
 17. Boud D. Assessment and learning- unlearning bad habits of assessment. Effective assessment at university conference, Queensland: Nov 1998
 18. 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 J Med Educ. 2013; 7(1):45-50.
 19. Liu, N. & Carless, D. Peer feedback: the learning element of peer assessment. Teaching in Higher Education. 2006; 11(3):279-290.
 20. 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.
 21. Menon AR, Kannambra PN. Item Analysis to Identify Quality Multiple Choice Questions. National Journal of Laboratory Medicine 2017 Apr; 6(2): MO07-MO10. Available on : DOI: 10.7860/NJLM/2017/25690:2215
 22. Mitra N K, Nagaraja H S, Ponnudurai G, Judson J P. The Levels of Difficulty And Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests. IeJSME 2009;3 (1): 2-7
 23. Meshkani Z, Abadie H. Multivariate analysis of factors influencing reliability of teacher made tests. Journal of Med Ed Winter. 2005; 6:149-159
 24. Marso RN, Pigge FL. An analysis of teacher made tests and item construction errors. J Contemp Edu Psych 1991; 16: 279-286.
 25. Mehrens, W. A., Lehman, I. J. Measurement and evaluation in education and psychology (4th edition). Belmont, CA: Wadsworth/Thomson Learning; 1991.
 26. Baner D, Kopp V, Fischer MR. Answer changing in multiple choice assessment change that answer when in doubt and spread the word. BMC Med Education. 2007; 7: 28.
 27. Meckenzie J. Vague and ambiguous questions on multiple choice exercises: the case for educational philosophy and theory. 1994; 26:23-33.
 28. Matlock-Hetzel, S. Basic concept in item and test analysis. A paper presented at annual meeting of the Southwest Educational Research Association. Austin, Texas: 1997. Available on: www.ericae.net/ft/tamu/espy.htm
 29. Kline T. J. B. Psychological testing: A practical approach to design and evaluation. Thousand Oaks, CA, US: Sage Publications Inc.; 2005.
 30. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc 2012; 62: 142-6.