International Journal of Clinical Biochemistry and Research 2019;6(3):330–335

Content available at: iponlinejournal.com

International Journal of Clinical Biochemistry and Research

Journal homepage: www.innovativepublication.com

Original Research Article

Item analysis of biochemistry multiple choice questions - A retrospective study & scope of improvement in method of assessment of MBBS students

Janki Desai[1],*, Vishal Shah[1]

[1] *Dept. of Biochemistry, GMERS Medical College & Hospital, Ahmedabad, Gujarat, India*

ARTICLE INFO

Article history:
Received 26-07-2019
Accepted 21-08-2019
Available online 21-09-2019

Keywords:
DE
DI
DIFI
MCQ
NFD

ABSTRACT

Introduction: Multiple choice questions (MCQs) are frequently used to assess in depth Knowledge of students in various subjects. In fact, a rightly frame MCQ with acceptable quality, measured by Difficulty Index (DIFI), Discrimination Index (DI) and Distractor Efficiency (DE) is the tool for assessment of student's performance.
Objective: To evaluate MCQs or items with their DIFI, DI and DE & to make database of valid items for future examination that increases creditability of Assessment.
Materials and Methods: MCQ examination was conducted as a part of Day to Day Assessment for 1st year MBBS students in Biochemistry Department at the end of 2nd term. 05 Set of 30 MCQs each had been randomly distributed among 126 students and each MCQ / Item was analyzed. Data was entered and analyzed in MS Excel 2007 for statistical analysis. Mean, Standard Deviations were calculated of DIFI, DE & DI for each items.
Results: Out of 30 items, 11 had "good to excellent" DIF I (31 - 60%) and 1 6 had good to excellent " DI ($>$ 0.25). Mean DE was 64.4% considered as nearer to acceptable. Out of all 05 groups & 90 D istractors average 30-32 were Non Functional Distractors. Negative DI indicates poor preparedness of students and issues with framing of some of the MCQs.
Conclusion: Study emphasizes, each assessment test of MCQs can be improved by developing improved items, maintaining database & selection of quality MCQs which truly assess the knowledge of students according to their abilities to differentiate in correct manner.

1. Introduction

Assessing the student performance is very important when learning goals involve the acquisition of skills that can be demonstrated by action.[1] One of the most important and authentic technique of assessing and estimating student performance across the full domain of learning outcomes as targeted by the instructors is classroom test.[2] Creating valid and reliable classroom test s are very important to instructor for assessing student performance, achievement and success in the class. One powerful technique available for guidance and improvement of In structor is Item Analysis.

Item analysis is a process which examines students responses to individual test items to assess quality of those items and test as a whole.[3] Item analysis helps Examiner to estimate the effectiveness of their test items i.e questions sets by involving various statistical techniques. Item analysis plays an important role in contributing to the quality of the test along with identifying content areas that maybe problematic for students and improving the item.

Multiple Choice Questions is one of the best tool for assessment of students. It requires thorough knowledge of subject, wide areas of content/subject can be enrolled during examination in less time period.[4]

A large proportion of curriculum is assessed in a short time requiring less effort of students but to make lot of effort and time spent by the examiners to make a quality MCQs.[2,5] Writing high quality distractors is an important part of the item and test development process.[2] Properly constructed

* *Corresponding author*.
E-mail address: drjankipatel315@yahoo.com (J. Desai).

https://doi.org/10.18231/j.ijcbr.2019.072
2394-6369/© 2019 Published by Innovative Publication.

multiple choice questions assess higher-order cognitive processing[2] of Bloom's taxonomy such as interpretation, synthesis and application of knowledge, instead of just testing recall of isolated facts.[5] In MCQs, each test item on a test is intended to sample student performance on a particular learning outcome. In recent trends having competitive examinations during various stages of students education i.e Entrance Examination for MBBS Course, Post Graduate Entrance Examination NEET-PG, various other Examinations like USMLE, GRE/ Toffel etc. for foreign Studies have similar examination pattern like MCQs.[6]

Graduate Entrance Examination NEET-PG, various other Examinations like USMLE, GRE/ Toffeletc for foreign Studies have similar examination pattern like MCQs.[6] All this is possible if the examiner knows the correct method of formulating a question. Single best answer MCQs consist of a question (the Stem), two or more options from which examinees select the one correct best response (the key) and the other option (the distractor).[2,5]

Thus standardizing the pattern for MCQs Questions having more reliable, valid test items along with involvement of Students in these examination pattern as their day to day assessment from their 1st Year of Graduate studies is one of the best tool to achieve highest competency in future studies.

Teaching by various methods will need a good/genuine assessment tool. By this study, our aim to get through a bridge connecting Teaching (to find problematic for students), Learning (To improve performance of students- Achievements) and Assessment (to develop reliable classroom Test) under the one roof.

Tests are too difficult (have insufficient floor) leads to frustration of the students and deflated scores whereas tests are too easy leads to decrease the motivation of the students and inflated scores.[7]

Tests can be improved in such a way that maintaining a pool of valid item for future use. In a question paper start with a easy question as a warm up questions to increase the moral of the students and difficult questions has put in between the easy questions so psychological boost up of the students and also identify top scorers.

Item analysis involves many statistical that can provide useful information for improving the quality and accuracy of MCQ and measure effectiveness of each test items. There are three main indices useful for analysis of items are: Difficulty index, Discrimination index and Distractor effectiveness. By using these indices the examiner is able to modify or remove specific items from the subsequent exams.

Item analysis also identifies the problematic area or misconception for the students about the particular concept and provides the opportunity to correct it.[7,8]

The quality of individual items is assessed by comparing students' item responsesto their total test scores. Reliability of the test also depends on the length and content of the items.[4]

2. Materials and Methods

126 students out of 150 1st Year MBBS students of GMERS Medical College & Hospital Sola were given test as their day to day assessment in Biochemistry Subject. The question paper was consist of 30 single best response type of MCQs (one mark each) in such a way that it includes almost all aspect of 1st Year MBBS Biochemistry Subject. All the quest ions had one correct answer with 03 distractors and students had given 01 hr time to answer all. 05 sets of Multiple Choice Questions paper had prepared. In each set of MCQ Paper, 20 MCQs are different & 10 Questions remain same but arrangement/sequence of MCQs had been changed to minimize errors. There was no negative Marking.

2.1. Data analysis

At the end of Examination the question papers had be en scored using the key and data was entered in the MS Excel 2007. Arrangement of the test papers in rank order from the highest score to the lowest score had b een done. For Item analysis the scored papers had arranged in order of marks obtained and then later had subdivided into 03 groups in higher ability group (H Group), Lower Ability Group (L Group) and middle group students. Set aside the Middle third (this was not be used in analysis). Frequency

Table by counting the number of students in the Higher Group and also for Lower Group who selected each alternative had been prepared. Based on these groupings, the following indices had calculated for all Items(questions) using standard methodology.[2]*(a) Difficulty Index* ; *(b)* Discrimination Index ; and *(c)* DistractorEfficiency (DE).[5]

Difficulty Index is an indicator of difficulty of the item(Questions) having inverse relationship. Lower the Difficulty Index, more difficult is the item. In simple terms, it tells the percentage of students in a group who could answer that item right.[1] It is also called facility value.[6]

Calculate item difficulty Index (P-Expressed in Percentage) by using formula:

P = (H+L) x100/T

Where H= no. of correct response in higher group

L= no. of correct responses in Lower group

T= total no. of responders from both the groups

Difficulty index ranges from 0 to 100%. If it is <30 (easy) or >70 (difficult) it is either revise or discard the question. Item difficulty is relevant for determining whether students have learned the concept being tested.[4] It is also between individuals' level of knowledge, ability, and preparedness.[7] The difficult items should be reviewed forpossible confusing language, areas of controversy oreven an incorrect key.[5]

Discrimination index is a measure of the ability of an item to differentiate between good and poor students.[6,7] It is ranges between -1 to +1. The closer the index is to +1, the more effectively the item distinguish between the two groups of students.[8] A negative discrimination index is most likely to occur with an item covers complex material written in such a way that it is possible to select the correct response without any real understanding of what is being assessed.[1] Sometimes an item will discriminate negatively,[9] such item should be revised and eliminated from scoring as it indicates that the lower performing students actually selected the key or correct response more frequently than top performers.[8] The poor students make a guess, select the response and come up with correct answer.[5] The good students may be suspicious of a question that looks too easy, may be harder path to solve the problem, read too much in to the question, and may end up being less successful than those who guess.[1] Such an item should be revised or discarded.

Calculate the discrimination index of the item (d) using the formula;

D= 2x (H-L)/T

If Discrimination index is between 0.15-0.24, item is good (store). If it is <0.15 item is poor so discard the item. If it is >0.2 5 item is excellent so store the item. Difficulty and discrimination indices are often reciprocally related.[5]

Distractor efficiency of each ite m is basis of the number of nonfunctional distractors in it and ranges from 0 to 100%. If an item contains three or two or one or nil NFDs then DE will be o, 33.3, 66.6 and 100% respectively.

Items were categorized as poor, good or excellent and actions such as discard/revise and store based on the value of DIF I and DI. In Distractor efficiency, how efficient each alternative was in distracting poor students but not good students. A good distractor should be picked up by at least 5% of the students.[1,2,5] If no one selects a distractor then it is revise the option to make it a plausible choice.[8]

2.2. *Study area*

Department of Biochemistry, GMERS Medical College & Hospital Sola

2.3. *Data collection method*

At the end of Examination, soft copy of all 30 Items with their choices and correct answer for all 150 students had been generated in Microsoft Excel. With taking help of Excel formulas Students had been divided in H & L group and various indices for Item Analysis have been calculated for every Item. Mean value of all indices was taken for getting any conclusion and discussion.

3. Result

Table 1 groups (30 MCQs in each group) and 90 distractors were analyzed. Score of 126 students ranges from 05 to 24 marks out of 30 marks question paper.

Table 1 represents the Mean and SD for DIF I, DI and DE were Group A has 52.5 ±26.0, 0.26±0.23 & 67.7±26.9, Group B has 51.1±25.3, 0.28±0.19 & 61.0±30.4, Group C has 48.5±27.7, 0.12±0.25 & 6 1.07±32.8, Group D has 52.0±19.1, 0.30±0.31 & 65.5±28.3, Group E has 50.6 ±20.7, 0.27±0.22 & 6 6.6±24.9, common group has 53.0±24.1, 0.63±0.23 & 93.3±14.0.

Table 2 represents interpretation and action taken according to the DIF I and DI.

If DIF I considered good to excellent and DI considered >0.25 only, items ideal for future storage were Group A had 11 items , Group B had 08 items, Group C had 09 items, Group D had 15 items, Group E had 12items and Common Group had 10 items.

Items had a negative and zero DI value in Group A had 05 items, Group B had 01 items, Group C had 13 items, Group D had 06 items, Group E had 6 items and common group had 00 items.

Table 3 represents Distractor analysis in which out of 90 distractors, average 30-35 distractors in each group were nonfunctional distractors. These 30-35 nonfunctional distractors in each group present in 20 items with DE varying from 33.3 to 66.6%. The 8-9 items had no NFD with their DE 100%

Table 4 represents relationship of Items with NFD and DIFI & DI. 5-11 Items of all groups have NFD showing Difficulty Index falls between 31-60 suggestive of good to excellent and 14-17 items showing DI more than 0.15 shows good discriminating power & storage of item can be done for further assessment.

4. Discussion

Single best correct response type MCQ is an efficient tool for evaluation. Its efficiency depends on the quality of MCQ which is assessed by the item and test analysis. Items having Difficulty Index value between 30-70 and DI ≥ 0.25 were considered as good difficulty and discrimination indices respectively. Each items must be evaluated based on Difficulty Index, DI and DE.

In this study, the mean Difficulty Index were in all Groups between 48-53%. These difficulty index were within the acceptable range of 31 to 60%. It means all MCQs had good and excellent Difficulty index so store the MCQs depend on the other indices. In other study of Mozaffer Rahim and Farhan Jaleel, the Difficulty Index value was 54.14± 17.48 which were also similar to my study.[5] Another Sakil study had 40-60% DIFI value.[9]

In another Mohammed AM study, the mean test scores ranged from 55.5% to 72%.[2]

Table 1: Assessment of 30 items based on various indices among Groups

	Group A		Group B		Group C		Group D		Group E		Group Common	
Parameter	**Mean**	**SD**	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Distractor Efficiency (DE%)	67.74	26.9	61.08	30.43	61.07	32.8	65.52	28.35	66.62	24.98	93.32	14.08
Difficulty Index (%)	52.54	26.0	51.14	25.36	48.54	27.7	52.08	19.16	50.68	20.73	53.04	24.19
Discrimination Index	0.26	0.23	0.28	0.19	0.12	0.25	0.30	0.31	0.27	0.22	0.63	0.23

Table 2: Assessment of 30 items based on DIFI & DI

DIFI	**Interpretation**	**Action**	**A**	**B**	**C**	**D**	**E**	**Common**
≤ 30	Difficult	Revise/Discard	6	8	8	4	6	2
31-40	Good	Store	7	5	4	4	3	1
41-60	Excellent	Store	4	3	5	11	9	3
≥ 60	Easy	Revise/Discard	13	14	11	11	12	4
Discrimination Index	Interpretation	Action	**A**	**B**	**C**	**D**	**E**	**Common**
<0.15	Poor	Discard/Revise	11	9	17	8	9	0
0.15-0.24	Good	Store	6	10	0	0	0	1
>0.25	Excellent	Store	13	11	13	22	21	9

Table 3: Assessment of 30 items based on distractor analysis

	Group A	**Group B**	**Group C**	**Group D**	**Group E**	**Common**
Total Distractors	90	90	90	90	90	30
Functional Distractors	61	59	55	59	51	28
Non Functional Distractors	29	31	35	31	39	2
Item with 1 NFD (DE-33.3%)	14	8	13	12	14	2
Item with 2 NFD (DE-66.6%)	6	12	5	8	11	0
Item with 0 NFD (DE-100%)	9	9	8	9	4	8

Table 4: Item with NFD and their relationship with DIFI & DI

Item with NFD and their relationship with DIFI						
	Group A	**Group B**	**Group C**	**Group D**	**Group E**	**Common**
<30	3	2	3	1	3	0
31-40	3	4	4	0	2	0
41-60	2	1	4	9	9	1
>60	12	13	7	10	11	1
Item with NFD and their relationship with DI						
	Group A	**Group B**	**Group C**	**Group D**	**Group E**	**Common**
<0.15	6	4	11	3	8	0
0.15-0.24	4	7	0	0	0	1
>02.5	10	9	7	17	17	1

In another Sarin study, most of the items had a low Facility Value, indicating that they were not correctly answered by majority of the students.[6]

In this study, the mean DI was in all groups between 0.26-0.63 in majority of groups except group C had less than 0.25. The DI value of >0.25 so the items were excellent DI and store in the MCQ bank for the future use. In Group C, DI was0.12 which is less than the acceptable limit of <0.15. This was happen because of in this group more easy questions had put and it should not able to discriminate the students. The item out of 30 items had DI less than<0.15 Group A had 11, Group B had 9, Group C had 17, Group D had 8, Group E had 9. In other study of Mozafferrahim and farhanjaleel,[5] the DI value of 0.356 ± 0.17. and had only 2/50 items with negative DI. The reason for the negative DI can be wrong key, ambiguous framing of questions. It means students of lower group were answered the questions more correctly than the higher group students. Another study of sarin[6] also had also DI value is far below the acceptable range.

In all groups had 18-20 items with 1 NFDS /2 NFDs had present except in common groups only 2 items with 01 NFDs are present it means majority of distractors are functional.

The non functional distractors is helpful in analyzing the usefulness of the items. The implausible distractors are removed or either revised according to the items difficulty and discrimination index. The designing of the plausible Distractors and reducing the NFDs is important aspect for framing the MCQs.

The difficulty index and the Discrimination index both are reciprocally related.[10] The questions with high DIF I easier questions, discriminate poorly and questions with low DIF I difficult questions are good discriminators.[8,9]

In this study, Mean DE of all groups were 64.4%. In other study of Mozafferrahim and farhanjaleel , the DE value of 81.4%.

In table 4 items with NFDs relationship with DIF I and DI which shows that in all groups except the Group E 05-15 items had a good to excellent difficulty and discrimination index although presence of NFDs. These NFDs are revised or re moved in the items for making the better construction of the items.

5. Conclusion

Developing the perfect test is the unattainable goal for anyone in an evaluative position. The item analysis is a simple procedure in evaluation. The reliability of the test measured by calculating the DIF I, DI and DE. The quality of MCQs frame by designing of plausible distractors and reducing non functional distractors. An ideal MCQ has Difficulty index between 31 to 60%, discrimination index >0.25 and DE 100% with three functional distractors.

In this study in all groups the difficulty index within the acceptable range so item analyzed were not too difficult or not too easy. Discrimination index in all groups except group C also within range >0.25 so these items were effectively discriminate the higher and lower group students. In group C has lower discrimination power <0.15 so did not discriminated the higher and lower group students. Items with negative DI and non functional. Distractor must be identified and revised or removed in the items for future examination.

While assessing the students by giving the different sets of papers the copying form the neighboring students should be avoided. Tests can be improved by developing improved items writing, better test design from which future tests can be drawn and that cover a reasonable span of difficulty levels.[7]

6. Source of funding

None.

7. Conflict of interest

None.

References

1. Matlock-Hetzel S. Basic concept in item and test analysis. Presented at annual meeting of the Southwest EducationalResearch Association. Austin ; 1997,. Available from: www.ericae.net/ft/tamu/espy.html.
2. Tarrant M, Ware J, Mohammed AM. An assessment of functioningand non-functioning distractors in multiple-choice questions:A descriptive analysis. *BMC Med Educ*. 2009;9:1–8.
3. Singh T, Gupta P, Singh D. Test and item analysis. In: Princ Med Educ ; 2009,. p. 70–77. 3rd ed.
4. Scorepak: Item analysis ; 2013,. Available from: Availablefrom:www.washington.edu./oea/score1/htm.
5. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficultyindex, discrimination index and distractor efficiency. *J Pak Med Assoc*. 2012;62:142–149.
6. Sarin YK, Khurana M, Natu MV, Thomas AG, Singh T. Item analysis of published MCQs. *Indian Pediatr*. 1998;35:1103–1105.
7. Eaves S, Erford B. The Gale group: The purpose of item analysis,item difficulty, discrimination index, characteristic curve ; 2013,. Available from: Availablefrom:www.education.com/reference/article/itemanalysis.
8. Scantron Guides - Item Analysis, adapted from Michigan StateUniversity website and Barbara gross devils tools for teaching ; 2013,. Last cited on. Available from: Availablefrom:www.freepdfdb.com/pdf/item-analysis-scantron.
9. Assessing student performance using test itemanalysis and its relevance to the State Exit Final Exams ofMAT0024 Classes-An Action Research Project from Mathematicsdept ;. Available from: Availablefrom:www.mdc.edu/hialeah/Polygon2008/articles/shakil3.pdf.
10. Sim SM, Rasiah RI. Relationship between item difficultyand discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Ann AcadMed Singapore*. 2006;35:67–71.

Author biography

Janki Desai Assistant Professor

Cite this article: Desai J, Shah V. Item analysis of biochemistry multiple choice questions - A retrospective study & scope of improvement in method of assessment of MBBS students. *Int J Clin Biochem Res* 2019;6(3):330-335.