Original Article

Indian Journal of Anatomy
Volume 13 Number 1, January - March 2024
DOI: http://dx.doi.org/10.21088/ija.2320.0022.13124.1

Distractor Efficacy, Difficulty and Discrimination Index of MCQ in Anatomy Formative Assessment: A Descriptive Analysis

Pushpalatha K[1], Sunil Kumar D[2], Manjunatha B[3], Deepa Bhat[4]

How to cite this article:

Pushpalatha K, Sunil Kumar D, Manjunatha B. Distractor Efficacy, Difficulty and Discrimination Index of MCQ in Anatomy Formative Assessment: A Descriptive Analysis. Ind Jr Anat. 2024;13(1):07-14.

Abstract

Background: Anatomy is the foundation stone of medical education, and assessing anatomy competency is essential. Assessment tools should fulfill objectivity, reliability, validity, and test higher cognitive skills. Assessment is a very significant component of education. To correctly judge the knowledge and skill of learners, assessment should be valid and should judge the appropriate levels of cognition. A large portion of the curriculum is assessed in a short period of time, requiring less effort from the student. However, it takes a lot of effort and time for the examiner to make high quality MCQs. Properly constructed multiple-choice questions assess higher-order cognitive processing of Bloom's taxonomy, such as interpretation, synthesis, and application of knowledge, instead of just testing recall of isolated facts

Methods: Hundred MCQ's from internal examinations of anatomy papers were analyzed for difficulty index, discrimination index (DI), and distractor efficiency (DE).

Results: In the present study, out of 100 MCQs, 19 were difficult, 60 were acceptable, and 21 were easy. 25 items did not discriminate against high achievers versus low achievers, 37 were acceptable, and 38 were highly discriminatory. There was a low positive correlation between the Difficulty Index and the Discrimination Index in tests two and three. However, there was a negligible correlation in tests one, four, and five.

Conclusion: Item analysis of the MCQs performed after the assessment provides insight into test item reliability and validity. In addition, it indicates how difficult or easy the questions were.

Keywords: Thermal burns; Feracrylum.

Author's Affiliation: [1]Professor, Department of Anatomy, [2]Professor & Head, Department of Community Medicine, [3]Professor, Department of Forensic Medicine, JSS Medical College, JSSAHER, Mysuru, Karnataka 570015, India.

Corresponding Author: Pushpalatha K, Professor, Department of Anatomy, JSS Medical College, JSSAHER, Mysuru, Karnataka 570015, India.

E-mail: pushpalathak@jssuni.edu.in

Received on: 23.01.2024

Accepted on: 20.02.2024

INTRODUCTION

Anatomy is the foundation of the medical sciences, and students need to attain core anatomical competencies for a strong basis for clinical correlation and professional practice.[1] Assessment is an essential component of teaching and learning. It is imperative to assess competence attained, so understanding the subject is a vital step towards performing and mastering the

competency.[2] Assessment in anatomy must obey the same general assessment parameters like objectivity, validity, and reliability.[1] The objectivity of assessment tools is important, irrespective of the purpose of the evaluation. Multiple-choice questions (MCQs) are one such tool that fulfills these parameters and is very commonly employed nowadays.

Designing MCQs is complex, difficult, and time consuming, but easy to administer. They are often used because of their high objectivity and ability to cover a wide area in a short amount of time. Higher levels of cognitive domains in Bloom's taxonomy can be assessed by properly constructed multiple-choice questions instead of just testing recall. This is done by formulating an item[3], and it needs to be tested for standard or quality.[4,2,5] Item analysis is a valuable yet relatively simple way of statistically analyzing assessment data. It is used to evaluate the quality and performance of test items performed after appearing in the examination.[3,2] The item analysis evaluates the difficulty level of the question and measures the ability of the item to discriminate exemplary students from others. The Difficulty index (DIF) and discrimination index (DI) are markers of the psychometric quality of the MCQ.[5]

The difficulty index is the fraction of students who get the answer correct. The higher the value, the easier the question, and it ranges from zero to 100. The higher the difficulty index value, the easier the questions. The discrimination index compares the performance of high achievers with that of low achievers, and the value may range from zero to one. If the discrimination index is higher, the item discriminates between high and low achievers. A third marker is distractor efficiency (DE). Percentage fluctuations in the distractors can be used to determine how well they distracted the students.[5-7] They can also be used to compare the quality of the distractors with the correct response. This helps to revise items and prepare for the test.[7] and indicates whether distractors function as distractors or not. If the distractors are selected by 5% or more of the students, they are considered functional distractors.[5]

Item analysis is critical to maintaining the quality of multiple-choice questions, and it is constructed from the responses given by students. The most difficult or easiest questions need to be revised or discarded, and they are based on the difficulty index, discriminatory index, and functional distractors.[8]

The present study evaluated the difficulty index, discriminatory index, and functional distractors of multiple-choice questions in a formative examination of anatomy.

METHODOLOGY

This was a descriptive longitudinal study. The analysis included 100 MCQs, 20 items per test, which were used in a formative assessment of anatomy for the 2019–20 and 2020–21 batches of medical students. Subject experts vetted and validated each of the MCQs based on the blueprint. 200 students evaluated each MCQ.

Data Collection: For item analysis, the results of all students were ranked in descending order, from highest marks to lowest marks, and then divided into quartiles. The top 27% scores as upper quartile or high scores (n = 54) and the bottom 27% scores as lower quartile or low scores (n = 54) groups were included in the analysis. The average scores, middle quartiles were excluded from the study.

Item analysis: Difficulty Index, discrimination index and distractor efficacy were calculated to evaluate MCQ.

Difficulty Index (DIF): $DIF = [(H+L)/N] \times 100$

H= Number of students who marked correct options in high score group

L=Number of students who marked correct options in low score group

T=Total number of students in both groups

DIF value is expressed in percentage. Its range is 0-100. Its recommended value is 45-60 and its acceptable value is 25-75[6]

Interpretation of DIF:

- DIF >70% = Too easy
- DIF b/w 30-70% = Acceptable
- DIF <30% = Too difficult

Discrimination Index (DI): $DI = 2\times[(H-L)/N] \times 100$

DI value is expressed in as a fraction. Its range is 0-1.0

Interpretation of DI:

- DI≤0.2 = Poor
- DI b/w 0.21-0.24 = Acceptable
- DI b/w 0.25-0.35 = Good
- DI≥0.36 = Excellent

Distractor efficacy (DE): Distractor efficiency was determined for each item based on the number of NFDs in it and ranged from 0 to 100%. DE was 100%, 66.6%, 33.3% and 0% based on presence of zero, one, two or three nonfunctional distractors in an item respectively.[3]

Distractor % = No of students selected the distractor/ Total no of students *100

Another way to analyze the distractors is calculating the item discriminator (ID).

ID= Nu-Nl/n

Nu = Number of students answered in upper group

Nl =Number of students answered in lower group

n= Number of students in 27%

Statistical analysis: All data are reported as mean ± SD of n items. The relationship between the item discrimination index and difficulty index values for each test paper was determined using regression analysis by statistical package of IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp and the coefficient of determination is given by R2. A *P* value of <0.05 was considered to indicate statistical significance.

RESULTS

A total of 200 students in each batch took the test. Three tests for 17-18 batch and two tests for 18-19 batch students were conducted and each test had 20 MCQ. A total of 100 Macq's and 300 distractors were analyzed.

Difficulty index (Table 1): Out of 100 questions, 19% were difficult, 21% easy and 60% are in the acceptable range. Test one has the highest number of easy questions followed by test five, then test three, and test two & fourhasthe least number of easy questions.

Discrimination index: (Table 2): Out of 100 questions 38% are excellent, 37% are acceptable and 25 % are poor in discriminating the good students from poor. Test three has the maximum number of excellent questions and poor questions, test four maximum number of acceptable & good questions.

Out of 100 items according to difficulty index 23 items need to be discarded and according to discrimination index 10 items need to be discarded.Testsone, three and five are more difficult than test two and four (Table 4).

There was a negligible correlation between the Difficulty Index and the Discrimination index (Fig. 1,

Table 1: Interpretation of Difficulty index for each test

Difficulty index Value	Interpretation	Action	Test 1 N (%)	Test 2 N (%)	Test 3 N (%)	Test 4 N (%)	Test 5 N (%)	Total N (%)
<30%	Difficult	Revise/ Discard	4 (20)	4 (20)	4 (20)	5 (25)	2 (20)	19 (19)
31-70	Acceptable	Store	9 (45)	14 (70)	12 (60)	13 (65)	12 (60)	60 (60)
>71	Easy	Revise/ Discard	7 (35)	2 (10)	4 (20)	2 (10)	6 (30)	21 (21)

Table 2: Interpretation of Discrimination index for each test

Discrimination index	Interpretation	Action	Test 1 N (%)	Test 2 N (%)	Test 3 N (%)	Test 4 N (%)	Test 5 N (%)	Total N (%)
<0.2 (Poor)	Poor	Revise/ Discard	5 (25)	6 (30)	7 (35)	4 (20)	3 (15)	25 (25)
0.21-0.24 (Acceptable)	Acceptable	Revise/ Discard	3 (15)	3 (15)	-	2 (10)	1 (5)	9 (9)
0.25-0.35 (Good)	Good	Store	5 (25)	3 (15)	3 (15)	9 (45)	8 (40)	28 (28)
>0.36 (Excellent)	Excellent	Store	7 (35)	8 (40)	10 (50)	5 (25)	8 (40)	38 (38)

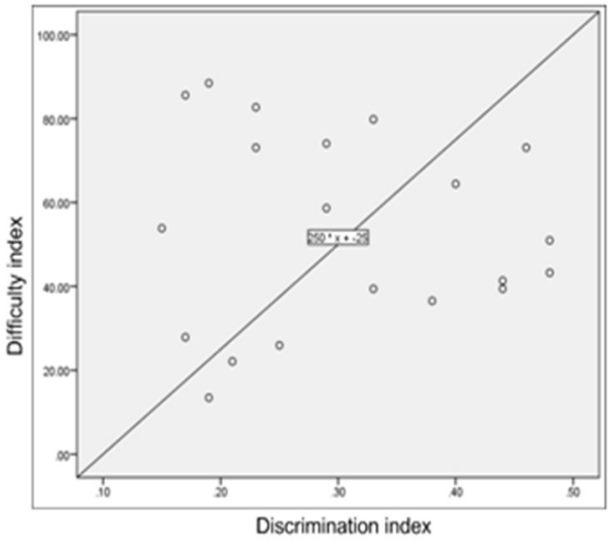

Fig. 1: Correlation between the Difficulty Index and the Discrimination index of test 1

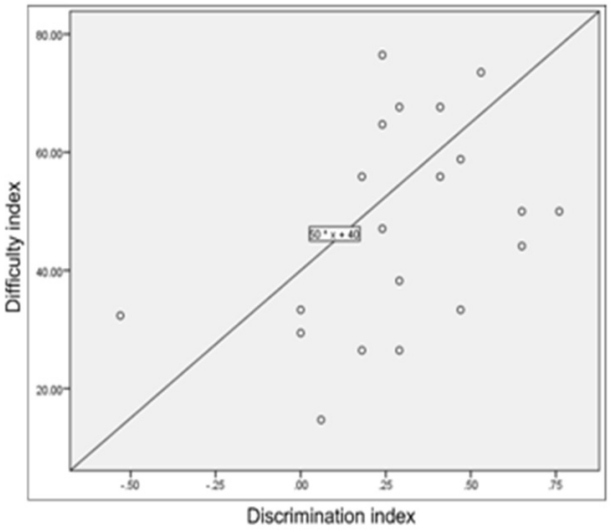

Fig. 2: Correlation between the Difficulty Index and the Discrimination index of test 2

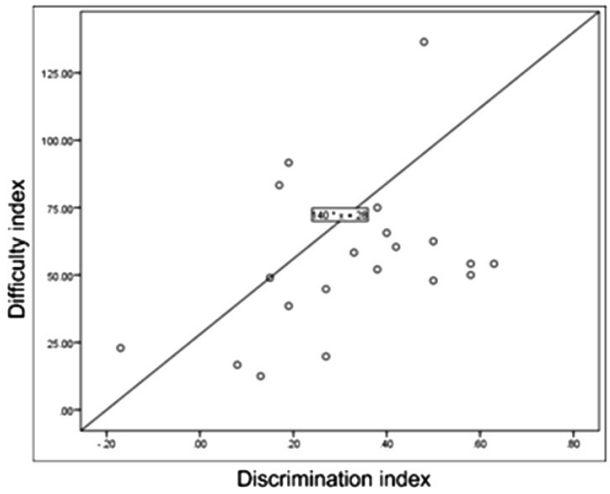

Fig. 3: Correlation between the Difficulty Index and the Discrimination index of test 3

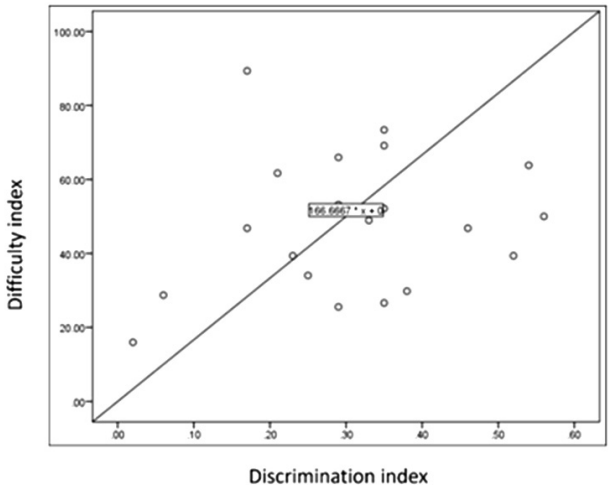

Fig. 4: Correlation between the Difficulty Index and the Discrimination index of test 4

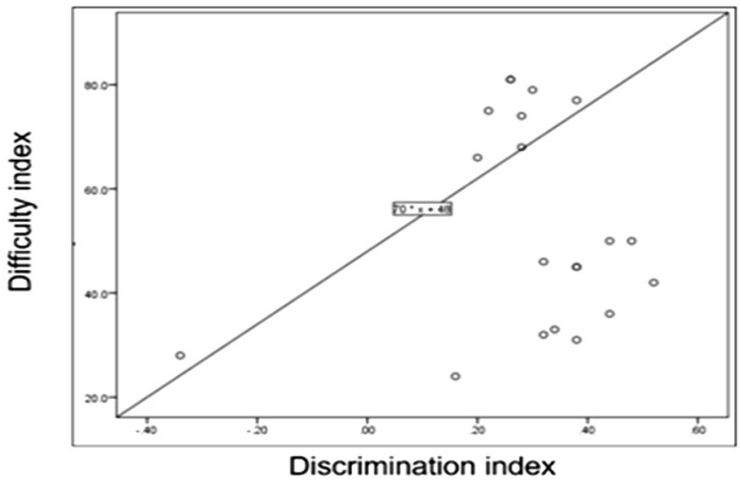

Fig. 5: Correlation between the Difficulty Index and the Discrimination index test 5

4&5) of Testone (r= -.051, p=.830), Testfour (r= .207, p=.381) and Test five. (r= .086, p=.717). It was observed that there was a low positive correlation between the Difficulty Index and the Discrimination index (Fig. 2&3) of Test two. (r= .400, p=.080) and Test three (r= .397, p=.083).

Table 3: Distractor efficacy

Number of NFD	Number of MCQ	Distractor Efficacy
0	63	**100%**
1	25	75%
2	11	50%
3	1	25%

Distractor efficacy (Table 3 & Fig. 6): Out of 300 distractors 250 were functional distractors and 50 were non functional distractors and 61 were not good distractors and need to be removed.

Distractors analyzed using item discrimination (ID), ID=Test two had a maximum number of non functional / not good distractors.

DISCUSSION

Assessment strategies are crucial for effective assessment, and their proper use presents a great challenge to educators. The assessment should be designed according to the objective. If the objective

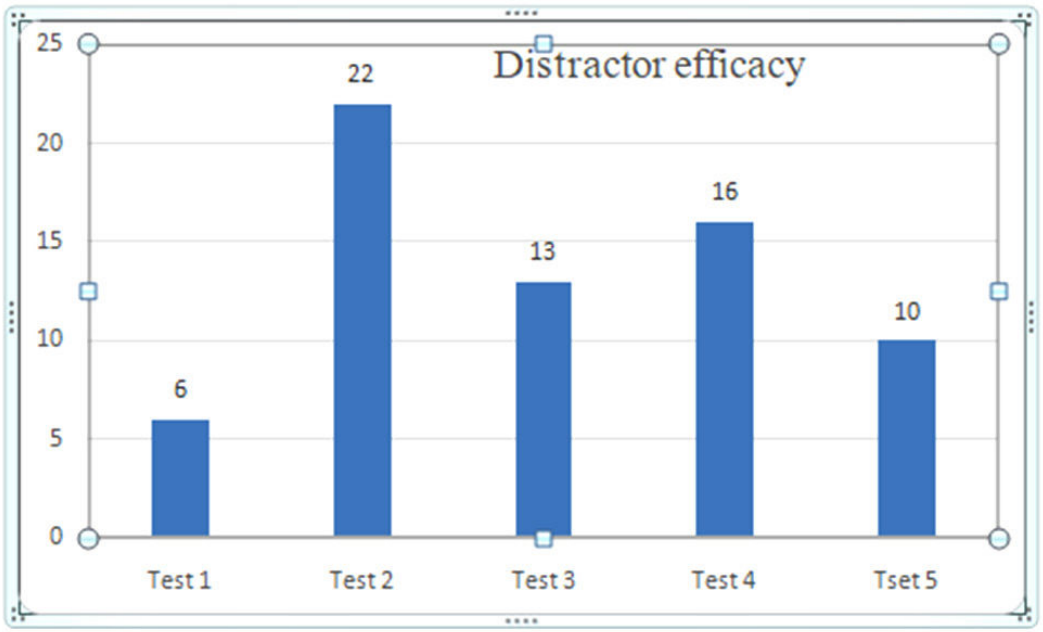

Fig. 6: Number of non functional/ not good distractors

Table 4: Mean and standard deviation of distractor efficacy, difficulty, and discrimination index of each test

Test	Difficulty index		Discrimination Index		Distractor efficacy	
	Mean	**St D**	**Mean**	**St D**	**Mean**	**St D**
1	53.702	23.027	0.306	0.114	-0.09	0.08
2	47.304	17.653	0.292	0.286	-0.1	0.14
3	54.793	28.500	0.323	0.201	-0.12	0.16
4	48.032	18.712	0.309	0.145	-0.08	0.1
5	53.15	19.914	0.3	0.177	-0.08	0.12

is not clear, we will fail to plan the assessment, leading to test failure.[9]

Several factors are considered for an effective assessment. Resource considerations are very significant, as this plays a major role in the exam chosen. The reliability and validity of an assessment play a significant role, and assessment tools should also be able to discriminate between a good and poor candidate. One such tool is MCQs, and they are highly reliable and valid. MCQ's are easy to mark and have a low impact on academic time compared to essay questions. However, its construction is a very time consuming and complex process.[10]

Item analysis assesses the quality of individual MCQ's and the whole test. It helps to identify topics that need more emphasis while teaching. The difficulty index and the discrimination index are often related.[11] Effective MCQs should have a proper difficulty level and discriminate between performers and non-performers. An item analysis helps to identify effective MCQ's depending on the difficulty index (DIF I), discriminating index (DI), and distractor efficiency (DE).[12,13]

The wide spread use of discrimination values with same level of difficulty may reflect some degree of guessing. Items with poor discrimination indices should be reviewed or discarded. This also gives feedback to the faculty regarding the test's validity. If the item is very difficult, it can be due tovariety of reasons. These include not being taught well, being inappropriate, or having the wrong wording or key given.[14]

In the present study, out of 100 MCQs, 19 were difficult, 48 were acceptable, and 33 were easy. Among the items, 25 did not discriminate against high achievers compared to low achievers, 37 were acceptable, and 38 were highly discriminatory. Out of 100 items, 77 could be retained, but 23 need to be discarded. There was a low positive correlation between the Difficulty Index and the Discrimination Index in tests two and three. However, there was a negligible correlation between tests one, four, and five. Tests one, three, and five were more difficult than tests two and four. This may be because the individuals responsible for creating the item were different.

It is possible that a successful student might not take the risk of attempting or taking a tough path to solving a difficult item. This might cause him to become unsuccessful. A weak student might guess and answer without understanding, and this could result in negative discrimination index.[14,11,15]

Distractors are vital components of an item and have a substantial impact on test score. The design of distractors influences student performance. A distractor is effective when more students from the group of low achievers choose it than from the group of high achievers. A distractor that is not chosen by students is dysfunctional, does not help measure educational outcomes, and has a negative impact on learners.[16]

For the items that have the highest number of non-distractors, it is inevitable that most students will get the items correct. If a high proportion of easy items are included in a test with absolute pass scores set at a fixed percentage (i.e., 50%), it will likely result in many borderline candidates passing who should not.[17]

In the present study, only 63 items had all distractors with 100% distractor efficacy, and only one item had all non functional distractors.

The number of non functional distractors affects an item's discriminative power. Items with one non functional distractor will have better discriminating abilities than items with four functioning distractors. Reducing the number of distractors from four to three decreases the difficulty index but increases the discrimination index and reliability. Because writing items with four distractors is difficult, the fourth distractor will mostly fill the gap and become the weakest distractor.[9,18,19,20] Two or more functional distractors per item are required to say the item is good quality.[21] Distractor efficiency can be related to Bloom's cognitive level. Items at the application level will have an increased number of efficient distractors than items at the knowledge level.[22,23]

Although MCQ items measure analyzing skills, they cannot provoke cognitive processes as effectively as essays. These essays are better suited to testing higher cognitive levels of Bloom's taxonomy. Answering essay questions requires students to recall information from the text, but answering MCQs relies on familiarity. So, it is expected to correlate positively with the extent of active processing a reader engages in during text comprehension while answering essays. However, objectivity is difficult. Therefore, it is better to have a combination of MCQ's and essays in a test. Without changes in curricular or teaching learning activities, qualitative changes in assessment methods can alter student behavior and improve academic performance. This should be considered when planning a competency based curriculum.[15]

Entrance and licensing examinations prefer MCQs because they are easy to grade and objective. Some educators argue that MCQs require only knowledge recall, but many authors argue that well constructed MCQs can test higher order thinking. However, constructing MCQs to assess higher order questions is difficult and time consuming, and documentation of the cognitive levels associated with each question is very essential when interpreting student performance.[24] Assessment and case based questioning drive learning. The assessment tool influences students' critical thinking approach to learning. MCQ's with good context richness aid in complex clinical thinking, while questions with poor context mostly test recall levels of knowledge.[25]

Psychometrics not only reflect the quality of the items but also the quality of the students, and therefore, they cannot be taken as absolute values. So expert review is essential despite acceptable psychometric performance.[21] The quality of item also depends on faculty who constructed it and Bloom's taxonomy. Teachers can make necessary modifications to MCQs, like editing or deleting them, depending on the analysis results.[26,27]

The difficulty and discrimination index of four option items are lower than those of five option items. Psychometric properties remain same for three-option, four-option, and five-option tests. It is true that reducing the number of options may make the exam easier. However, removing the least distracting component is a vital strategy to maintain item's quality. It also reduces the time to answer questions, more content can be covered at the same time, and burden on the item writer is minimized.[28]

CONCLUSION

Item analysis is performed after the examination and provides information regarding test item reliability and validity. It also tells you how difficult or easy the questions were. Item analyses help detect specific technical flaws in the questions and provide information for improvement. Distractor efficiency is one such tool that tells whether the item was well constructed or failed to perform its purpose. Increases the examiner's item writing skill. Items having average difficulty and high discrimination against functioning distractors should be incorporated into future tests to improve test development and review. This would also improve the overall test score and provide proper differentiation among students.

Limitations: Limitation of this study was, different set of faculties were involved in preparing the items. There was no comparison between the different sets of students who were evaluated.

REFERENCES

1. Singh K, Bharatha A, Sa B, Adams OP, Majumder MA. Teaching anatomy using an active and engaging learning strategy. *BMC Med Educ.*2019; 19:149.https://doi.org/10.1186/s12909-019-1590-2
2. Manoor Narasimha Sachidananda Adiga, Swathi Acharya, Rajendra Holla. Item Analysis of Multiple-Choice Questions in Pharmacology in an Indian Medical School.*Journal of Health and Allied Sciences NU*. 2021; 11(03): 130-135.
3. DOI: 10.1055/s-0041-1722822
4. Chakravarty M, Latif NA, Abu-Hijleh MF, Osman M, Dharap AS, Ganguly PK. Assessment of anatomy in a problem-based medical curriculum. *Clin Anat*. 2005; 18:131-6. DOI: 10.1002/ca.20038
5. Brenner E, Chirculescu AR, Reblet C, Smith C. Assessment in anatomy. Eur J Anat 2015; 19:105-24.
6. Yaqinuddin A, Zafar M, Ikram MF, Ganguly P. What is an objective structured practical examination in anatomy? *Anat Sci Educ.*2013; 6:125-33.
7. Khan GN, Ishrat N, Khan AQ. Using item analysis on essay types questions given in summative examination of medical college students: Facility value, discrimination index. *Int J Res Med Sci.*2015; 3:178-82. DOI: 10.5455/2320-6012.ijrms20150131
8. 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. DOI: 10.4103/2395-2296.189670
9. Chauhan PR, Ratrhod SP, Chauhan BR, Chauhan GR, Adhvaryu A, Chauhan AP. Study of difficulty level and discriminating index of stem type multiple choice questions of anatomy in Rajkot. *Biomirror.*2013; 4:1-4.
10. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc*. 2012 Feb;62(2):142-7. PMID: 22755376..
11. Palmer, E.J., Devitt, P.G. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ*. 2007;7(1): 1-7.https://doi.org/10.1186/1472-6920-7-49
12. Mehta G, Mokhasi V. Item analysis of multiple choice questions- an assessment of the assessment tool. *Int J Health Sci Res*. 2014;4(7):197-202.
13. Juliana Linnette D'sa, Maria Liza Visbal- Dionaldo. Analysis of Multiple choice questions: Item difficulty, Discrimination Index and distractor efficacy. *International Journal of Nursing Education.* 2017;9(3):109-14. DOI:

10.5958/0974-9357.2017.00079.4

14. Wajiha Mahjabeen, Saeed Alam, Usman Hassan, Tahira Zafar, Rubab Butt, Sadaf Konain, Myedah Rizvi. Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Ann. Pak. Inst. Med. Sci.* 2017; 13(4);310-15. https://doi.org/10.48036/apims.v13i4.9
15. 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 Singap*. 2006 Feb;35(2):67-71. PMID: 16565756
16. Chery A. Melovitz Vasan, David O. DeFouw, Bart K. Holland,Nagaswami S. Vasan. Analysis of Testing with Multiple Choice Versus Open-Ended Questions: Outcome-Based Observations in an Anatomy Course. *Anat Sci Educ*. 2018; 11:254–61. doi: 10.1002/ase.1739.
17. Sharma L R. Analysis of Difficulty Index, Discrimination Index and Distractor Efficiency of Multiple-Choice Questions of Speech Sounds of English. *International Research Journal of MMC*. 2021; 2;1. DOI: https://doi.org/10.3126/irjmmc.v2i1.35126
18. Tarrant M, Ware J andMohammed A M. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education.* 2009; 9:40. https://doi.org/10.1186/1472-6920-9-40
19. Sahoo DP, Singh R. Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. *Int J Res Med Sci*. 2017; 5:5351-5.DOI: https://doi.org/10.18203/2320-6012.ijrms20175453
20. Ansari M, Sadaf R, Akbar A, Rehman S, Chaudhry ZR, Shakir S. Assessment of distractor efficiency of MCQS in item analysis. *Professional Med J*. 2022; 29(5):730-34. DOI: https://doi.org/10.29309/TPMJ/2022.29.05.6955
21. Date AP, Borkar AS, Badwaik RT, Siddiqui RA, Shende TR, Dashputra AV. Item analysis as tool to validate multiple choice question bank in pharmacology. *Int J Basic Clin Pharmacol*. 2019; 8:1999-2003.DOI: https://doi.org/10.18203/2319-2003.ijbcp20194106
22. Thomas Puthiaparampil and Mizanur Rahmanz.. How important is distractor efficiency forgrading Best Answer Questions? *BMC Medical Education*. 2021; 29:2- 6. https://doi.org/10.1186/s12909-020-02463-0
23. Testa S, Toscano A and Rosato R. Distractor Efficiency in an Item Pool for a Statistics Classroom Exam: Assessing Its Relation with Item Cognitive Level Classified According to Bloom's Taxonomy. *Front. Psychol*.2018; 9:1585.https://doi.org/10.3389/fpsyg.2018.01585
24. Sajjad M, Iltaf S, Khan RA. Nonfunctional distractor analysis: An indicator for quality of Multiple-choice questions. *Pak J Med Sci*. 2020;36(5):982-86. doi: 10.12669/pjms.36.5.2439
25. Andrew R. Thompson, Valerie D. O'Loughlin. The Blooming Anatomy Tool (BAT):A Discipline-Specific Rubric for Utilizing Bloom's Taxonomy in the Design and Evaluation of Assessments in the Anatomical Sciences. *Anat Sci Educ.* 2015; 8:493–501.
26. DOI: 10.1002/ase.1507
27. Magalie Chéron, Mirlinda Ademi, Felix Kraft and Henriette Löffler-Stastka1Case-based learning and multiple-choice questioning methods favored by students. *BMC Medical Education.*2016;16:41. https://doi.org/10.1186/s12909-016-0564-x
28. Patil R, Palve SB, Vell K, Boratne AV. 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.DOI: https://doi.org/10.18203/2394-6040.ijcmph20161638
29. Burud I, Nagandla K, Agarwal P. Impact of distractors in item analysis of multiple-choice questions. *Int J Res Med Sci.* 2019; 7:1136-9.DOI: https://doi.org/10.18203/2320-6012.ijrms20191313
30. Nicolette Fozzard, Andrew Pearson, Eugene du Toit, Helen Naug, William Wen and Ian R. Peak. Analysis of MCQ and distractor use in a large first year Health Faculty Foundation Program: assessing the effects of changing from five to four options. *BMC Medical Education.* 2018; 18:252. https://doi.org/10.1186/s12909-018-1346-4