

EVALUATION OF THE MULTIPLE-CHOICE QUESTION ITEM ANALYSIS OF THE SIXTH YEAR UNDERGRADUATE ORTHODONTIC TESTS AT THE FACULTY OF DENTISTRY, KING ABDULAZIZ UNIVERSITY, SAUDI ARABIA

Fadia M. Al-Hummayani, DDS, MSc.¹

Abstract

Introduction: Poorly constructed multiple-choice questions (MCQs) are still being reused in medical and dental colleges. Item analysis reports are used to evaluate MCQ tests and eliminate the substandard MCQs for quality improvement purposes.

Aim: This study aims to assess the quality of the MCQ item analysis of dental students' final year orthodontics examination using the difficulty index (DIF I), discrimination index (DI), and distractor efficiency (DE) parameters. Additionally, the relationship between these parameters is determined to categorize the tested MCQs into either acceptable, which would be stored in the MCQ bank, or substandard (poor), which would be eliminated or rephrased.

Methods: Four MCQ examinations from the sixth-year undergraduate orthodontic course of the academic year 2018–2019 were evaluated by item analysis using the DIF I, DI, and DE. A total of 165 MCQ items (495 distractors) were collated and analyzed after 189 students answered the tests.

Results: The mean and standard deviation of the DIF I and DI were 0.76 (± 0.19) and 0.25 (± 0.14), respectively. Distractor assessments revealed that 27.5% were non-functional distractors (NFDs) with a DE of 78.2%. There was a significant negative correlation between the DIF I and the DI ($r: -0.711$; $p < 0.01$). A significant positive correlation was found

between the DIF I and NFDs ($r: 0.644$; $p < 0.01$). Of the 165 MCQ items, 153 were considered sufficiently good to be stored in the MCQ bank, and only 12 were poorly constructed and recommended for elimination or rephrasing.

Conclusion: Item analysis was effective in identifying the good and poorly constructed MCQ items. Therefore, it should be used with each university MCQ examination for the development of a reliable and valid MCQ bank.

Keywords: Item Analysis, Difficulty Index, Discrimination Index, Non-Functional Distractors).

Introduction

Multiple Choice Question (MCQ) examinations are extensively used as an educational examination tool in many institutes.¹ Many believe that a well-constructed MCQ test is an unbiased assessment that can measure knowledge and is necessary to reflect students' performance in a course.^{2,3} Therefore, high-quality MCQ tests are essential for accurate education level assessment.⁴

Item analysis is used to evaluate MCQ tests by examining the responses of the student to individual MCQ test items using distractor efficiency (DE) and the difficulty and discrimination item indices.⁵ The difficulty index (DIF I) is useful for analysis. It

¹ Dr. Fadia M. Al-Hummayani Assistance Professor at the Department of Orthodontics, Faculty of Dentistry, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia. E-mail: falhummayani@kau.edu.sa

measures how many exam-takers answered the MCQ item accurately. The values range from 0 to 1. A low DIF I indicate that the MCQ item is very difficult and not easy to comprehend. It is desirable for the MCQ item to have a moderately difficult index score (> 0.3 and < 0.8).⁶ The discrimination index (DI) is valuable for analysis. It determines if the correct answer option for the MCQ item is effective in distinguishing between the higher-performing knowledgeable students and the lower-performing students who are not as knowledgeable. Its values range from -1 to +1. If the correct answer option for the MCQ item has a negative DI, it indicates that the reliability of the test is low and that it should be eliminated.^{7,8,9}

Distractor analysis is used to identify the functional distractors (FDs) and the non-functional distractors (NFDs) in an MCQ item. An effective distractor is a FD because it is similar to the correct answer. Hence, an FD should indicate what the students commonly misunderstand regarding the correct answer. It can be identified by its frequent selection by exam-takers ($> 5\%$) as the correct answer. An NFD, however, is a poorly constructed and ineffective distractor. It is identified by its infrequent selection by exam-takers ($< 5\%$). It is termed a poorly constructed distractor because it is either not a true statement, or it is unrelated to the question being asked. DE is calculated based on the number of NFDs for the MCQ item. Its value ranges from 0% to 100%; if the number of NFDs for the MCQ item is zero, the DE is 100%.^{10,11}

Some researchers have studied the relationship between MCQ item difficulty Index (DIF I) and the discrimination power (DI) of the correct answer option and found an indirect relationship; for an MCQ with a high DIF I (easy MCQ), the DI of the correct answer

option is low. The relationship between MCQ item DIF I and the number of NFDs has also been studied and a direct relationship has been found; as DIF I increases (easy MCQ), the number of NFDs increases.¹²⁻¹⁵

At the faculty of dentistry in King Abdulaziz University (KAU), Jeddah, Saudi Arabia, the examination committee weighed the advantages and limitations of multiple-choice tests and decided to use them in all the undergraduate university examinations. Few studies on MCQ tests' item analysis were conducted at King Abdulaziz University in the faculty of medicine. However, to our knowledge, no studies were conducted at the faculty of dentistry at the same university (KAU). Although all the departments of the college of dentistry at KAU routinely conduct item analysis for all the MCQ exams, the importance of the item analysis reports is underestimated by the faculty members who create the MCQ tests, and the reports on MCQ item analysis feedback are inadequate. In fact, the automatically generated analysis is not useful, and the same MCQ mistakes are repeated.^{1,2,3} Therefore, for quality improvement purposes, this study aimed to assess the item analysis of four MCQ examinations, namely the dental students' final year "Orthodontics" examinations. This assessment was conducted using the DIF I the DI, and the DE. The relationship between them was also determined to categorize the tested MCQs (depending on the values of DIF I, DI, and the number of NFDs). Thus, the MCQs were categorized as either acceptable, to be stored in the MCQ bank, or substandard (poor), to be eliminated or rephrased.

Materials and Method

The Research Ethics Committee at King Abdulaziz University in Jeddah, Saudi Arabia,

approved this cross-sectional descriptive study, RCE # 049-02-19.

DATA COLLECTION AND ANALYSIS

Item analysis reports of four MCQ examinations for the sixth-year undergraduate orthodontic course of the academic year 2018–2019 were used. These MCQ examinations were formulated by 14 orthodontic staff members teaching the sixth-year orthodontic course at the faculty of dentistry, KAU. A total of 189 students took the sixth-year undergraduate orthodontic course and answered these four MCQ tests “Hence, they are the MCQ test-takers”. Each MCQ item included a stem and four options: three distractors and one correct answer. One mark was awarded for the correct answer, and zero was awarded for the incorrect response. Negative marking was not used.

The four MCQ examinations were two quizzes (each examination quiz contained 10 MCQ items), the midterm examination (50 MCQ items), and the final examination (95 MCQ items). From these tests, 165 MCQ items with 660 options (165 correct answers and 495 distractors) were collated. The MCQ tests were assessed for difficulty level using the DIF I of

each MCQ item and for discrimination power between low-achieving and high-achieving students using the DI of the correct answer for each MCQ item. Finally, a distraction assessment of the three distractors in each MCQ item was performed to identify the FDs and NFDs to analyze the DE.

The DIF I was computed by means of the formula $H+L/N \times 100$, where H and L are 26% of the high-performing and low-performing students with the correct responses, and N is the total number of exam-takers in each group. The DIF I value ranged between 0 and 1. High values indicated a low difficulty level (easy MCQ item) and vice versa. Furthermore, the DI was computed to evaluate the correct answer option’s discrimination power between the high-achieving and low-achieving student groups using the formula $H-LX2/N$. It ranged between -1 and +1. High values indicated a higher discrimination power and vice versa. Based on Ebel and Frisbe’s (1972) guidelines on classical test theory item analysis as explained by *Haladyna and Downing (1989)*.¹⁶ The DIF I and DI were categorized as seen in (Table1).

Table 1. Ebel and Frisbe’s (1972) interpretation of difficulty and discrimination indices

Difficulty Index (DIF I)		Discrimination Index (DI)	
Difficult	0 to 0.30	Poor	Negative to 0
Moderately Difficult	0.31 to 0.79	Fair	0.01 to 0.19
Easy	≥ 0.80 to 1	Good	0.20 to 0.39
		Excellent	≥ 0.40 to 1

DIF I value > 0.30 indicate acceptable moderately difficult MCQ item, and values ≥ 0.8 indicate an easy MCQ item (low difficulty level). However, the MCQ item is only considered acceptable if the distractors are effective. Values lower than 0.31 indicated very difficult MCQ items, and these items were recommended to be removed or rephrased. Regarding the discrimination power of the correct answer option in an MCQ item using the DI, an acceptable DI value was between 0.01 and 1, and a DI value ≥ 0.4 was considered excellent. A negative DI value (less than zero) was considered unacceptable. Distractor assessment was performed by identifying the FDs and NFDs for each MCQ item. If the distractor was chosen by less than 5% of the exam-takers, it was considered an NFD, and if the distractor was chosen by more than 5% of the examinees, it was considered an FD. From the distractor analysis, the DE was calculated based on the number of NFDs in the MCQ item. The DE ranged from 0% to 100%. If the MCQ item had no NFDs, the DE was 100%, and if the MCQ item had one, two, or three NFDs, the DE percentages of the MCQ item were 66.66%, 33.33%, and 0%, respectively. The relationship between the DIF I, DI, and the number of NFDs was calculated using a two-tailed ANOVA Pearson's correlation test.

It was essential to categorize the MCQ items based on their item analysis results to update the MCQ bank with reliable and valid MCQ items. Therefore, based on the results of the DIF I, DI, and NFDs of the four MCQ tests in this study, the MCQ items were categorized as follows:

- MCQ items with a moderate difficulty level (DIF I between 0.31 and 0.79), high

discrimination power of key answer (DI between 0.20 and 1), and zero NFDs were considered "Ideal MCQ Items" and recommended to be used frequently in future examinations because they increased the MCQ test's reliability.

- MCQ items with an easy difficulty level (DIF I between 0.8 and 1), fair to high discrimination power of the key answer (DI between 0.01 and 1), and one or two NFDs were considered easy but acceptable MCQ items, classified as "Good MCQ Items," and recommended to be stored in the MCQ bank to be reused in future examinations.

- MCQ items with an extremely high difficulty level (DIF I between 0 and 0.30) "low discrimination power of the key answer (a negative DI value or zero) with three NFDs were considered very difficult, misleading and useless, classified as "Poor MCQ Items," and recommended to be eliminated or rephrased before being reused.

Results

The average examination score of all four MCQ tests was 86%. The frequencies and percentage values of the difficulty and discrimination indices of the 165 MCQ items, as suggested by Ebel and Frisbe's (1972)³ test theory of item analysis are presented in Table 2. The difficulty levels of the 165 MCQs of the four tests were classified as moderately difficult (47.9%), easy (48.5%), and very difficult (3.6%). The mean DIF I was 0.76 (SD ± 0.19). Further, more than 50% of the correct answer options showed fair discrimination power, and 41.8% had good to excellent discrimination power. The mean and standard deviation of the DI was 0.25 (SD ± 0.14). A bar chart representation of the DI percentages of the MCQ answer key is shown in Figure (1).

Table 2. Difficulty index and discrimination index frequencies and percentage values of each MCQ test with the overall mean and SD of the DIF I and DI

MCQ Examinations n (# of MCQs)	Difficulty Index (DIF I) n (%)			Discrimination Index (DI) n (%)				Score %
	Difficult (0 – 0.30)	Moderate Dif (0.31 – 0.79)	Easy (≥ 0.8 to 1)	Poor (negative)	Fair (0.01 – 0.19)	Good (0.2 - 0.39)	Excellent (≥ 0.4 – 1)	
Quiz 1 (10)	1 (10%)	8 (80%)	1 (10%)	0 (0%)	3 (30%)	3 (30%)	4 (40%)	78%
Quiz 2 (10)	1 (10%)	6 (60%)	3 (30%)	0 (0%)	3 (30%)	4 (40%)	3 (30%)	80%
Midterm (50)	0 (0%)	26 (52%)	24 (48%)	2 (4%)	32 (64%)	9 (18%)	7 (14%)	90%
Final (95)	4 (4.2%)	39 (41%)	52 (54.8%)	10 (10.5%)	46 (48.4%)	32 (33.7%)	7 (7.4%)	96%
Total (165)	6 (3.6%)	79 (47.9%)	80 (48.5%)	12 (7.3%)	84 (50.9%)	48 (29.1%)	21 (12.7%)	
Mean ±SD (%)	0.76 (±0.19) = 76% (±19%)			0.25 (±0.14) = 25% (+14%)				86%

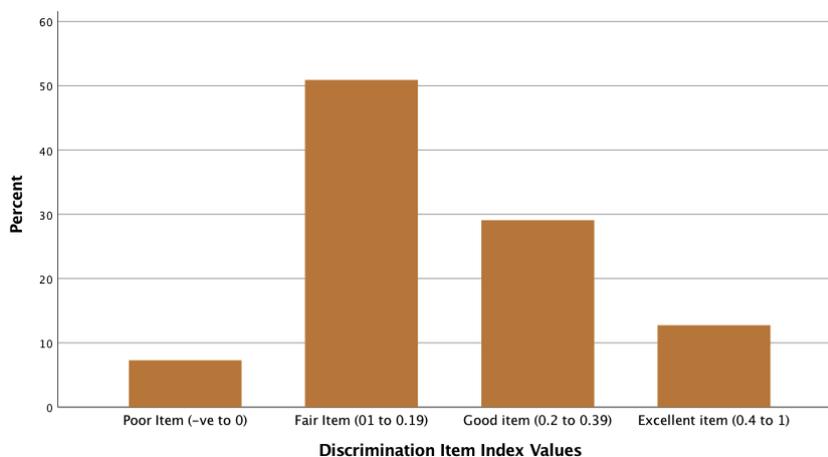


Figure 1. Bar chart of the discrimination index percentages of the MCQ answer key

Table 3 shows the result of the discrimination assessment. Of the 495 distractors, 359 (72.5%) were found to be FDs, and 136 (27.5%) were found to be NFDs.

Table 3. The overall frequency and percentage of the non-functional distractors (NFDs) and functional distractors (FDs)

Distractor Type	Frequency	Percentage (%)
Functional Distractors (FDs)	359	72.5%
Non-Functional Distractors (NFDs)	136	27.5%
Total	495	100%

The distribution of NFDs per MCQ item is shown in Table 4. Of the 165 MCQ items, only seven MCQ items had three NFDs. However, 73 MCQ items had zero NFDs, and the remaining MCQ items (85 MCQ items) had

one or two NFDs. Hence, 44.2% of the MCQ items had effective distractors (100% DE), and only 4.2% of the MCQ items exhibited 0% ineffective distractors.

Table 4. Frequency and percentage of NFDs per MCQ item

Number of NFDs	Distractor Efficiency (DE)	Frequency	Percentage (%)
Zero	100%	73	44.2%
One	66.66%	55	33.3%
Two	33.33%	30	18.2%
Three	0%	7	4.2%
Total	-----	165	100%

Difficulty and discrimination indices correlation

The two-tailed ANOVA Pearson’s correlation between DIF I and DI showed a significant negative correlation (r: -0.712; p < 0.01).

Correlation between the difficulty index and the number of non-functional distractors

The ANOVA test showed a significant positive correlation (r: 0.644; p < 0.01) between the DIF I and the number of NFDs.

The frequency and percentage of the 165 MCQ items categorized as ideal, good, or substandard (poor) MCQ items according to their DIF I, DI, and NFDs are presented in Table 5 and represented in bar chart Figure (2). Of the 165 MCQ items, 12.7 % were “Ideal MCQ Items,” and 80% were “Good MCQ Items.” Both were recommended to be stored in the MCQ bank to be reused for future tests. However, only 12 MCQ items (7.3%) were considered “Poor MCQ Items” and recommended to be eliminated or rephrased.

Table 5. Frequency and percentage of MCQ type (substandard, good, and ideal) distributed according to DIF I and DI values and number of NFDs

MCQ Type and Fate Decision	Frequency	Percentage (%)
Substandard (poor) MCQ to be removed DIF I = (0 to 0.3), DI = (negative), NFDs = 3	12	7.3%
Good MCQ to be stored in bank DIF I = (0.31 to 0.79), DI = (0.01 to 0.39), NFDs =1 or 2	132	80%
Ideal MCQ to be stored in bank DIF I = (0.31 to 0.79), DI = (0.2 to 0.39), NFDs = 0	21	12.7%
Total	165	100%

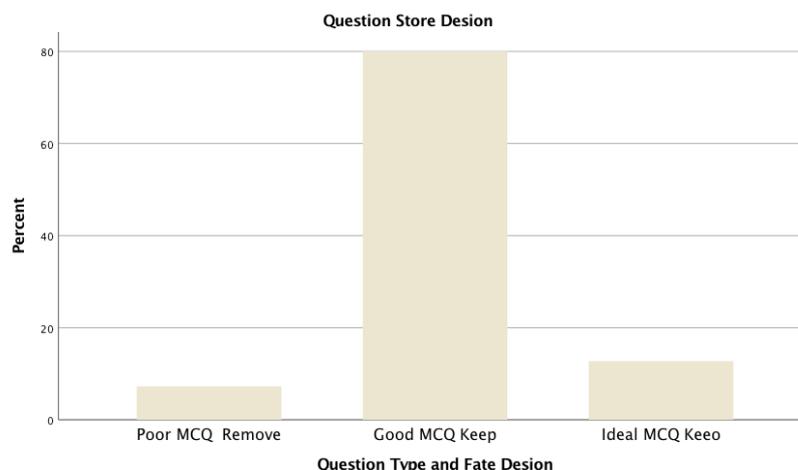


Figure 2. Percentages of the MCQ type (substandard, good, and ideal)

Discussion

Examinations using multiple-choice questions have become one of the main assessments at universities worldwide. In King Abdulaziz University (KAU), Jeddah, Saudi Arabia, MCQ tests are extensively used for undergraduate examinations for two reasons. First, because the examination committee believes that, compared with other tests, a well-constructed MCQ test evaluates a wide range of higher-order thinking skills quickly and efficiently. Second, because MCQ examinations provide a quick and easy method of scoring by hand or electronically. This is essential because the increasing number of undergraduate students enrolling in the university makes it difficult for educators to submit examination marks in a limited time frame. Hence, evaluating the item analysis report of every MCQ test is essential to ensure high-quality MCQ examinations.^{1,2,3}

In this study, the 165 evaluated MCQ items had a DIF I percentage mean of 76% and an SD of $\pm 19\%$, indicating an acceptable difficulty level. The DI mean was 0.25, with an SD of +0.14, indicating good discrimination power of the key answer. These results were similar to results from other institutions, such as the study on the MCQ tests at the pediatric department of the Arabian Gulf University, Manama, in which the mean DIF I was 73% and the DI was 0.20.⁷ In addition, *Omer AA et al. (2016)*, reported similar findings, with a mean DIF I of 74% and a DI of 0.20 in a study on a surgical examination administered to 44 sixth year medical students at the faculty of medicine, University of Tabuk, Saudi Arabia.¹⁷ Additionally, in a study at Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE, conducted to assess the effectiveness of multiple-choice question items in two postgraduate pediatric dentistry tests,

the DIF I mean was 89% and 76%, and the DI mean was 0.25 and 0.32, respectively.¹⁸ *Date AP et al. (2019)*, conducted an item analysis evaluation of MCQ tests of the pharmacology department at Salve Institute of Medical Sciences, India, and found the mean DIF I to be 70%, with a DI of 0.26.¹⁹ Hence, the majority of the studies including the present study agree that a moderately difficult MCQ item with a good DI key answer is an effective examination tool.^{15, 17, 18, 19}

In our study, the correlation between difficulty and discrimination indices showed a significant negative correlation ($r: -0.712$; $p < 0.01$), indicating an indirect relation; the DI of an MCQ item with a high DIF I value (easy MCQ) is low. Similar observations were reported in many studies such as *Rao C et al. (2016)*, *Gajjar S (2014)*, *Mitra et al. (2009)*, and *Si-Mui Sim et al. (2006)*. These studies showed that the DI was negatively correlated with the DIF I, indicating that a very easy MCQ would likely have a low level of discrimination. This is usually because low-performing students are as likely to choose the correct answer as high-performers with these MCQ items, resulting in the MCQ item being useless due to its poor DI.^{13,14,21,22}

In addition, this study found a significant positive correlation ($r: 0.644$; $p < 0.01$) between the DIF I and the number of NFDs, indicating that as the DIF I increased (easy MCQ), the number of NFDs increased. Hence, as the number of NFDs increases, the DE of the MCQ item decreases. This finding aligns with the findings of *Date AP et al. (2019)* and *Mehta & Mokhasi (2014)*.^{9, 19}

Student achievement is related to the construction of distractors. Distractor analysis detects inaccuracies requiring review, substitution, or elimination.^{13, 14} Our analysis demonstrates that from the 495 distractors, 359

(72.5%) were FDs and 136 (27.5%) were NFDs and the DE percentage of the 165 MCQs ranged from 100% to 66%. *Date AP et al. (2019)*, showed similar findings after testing 120 distractors. Of them, 84 (70%) were (FDs), and 36 (30%) were NFDs. with DE of 66% to 100%.¹⁹

When creating an MCQ bank, detection of very difficult and very easy MCQ items is crucial, and the elimination or rephrasing of such MCQ items is required before reusing them in future examinations. The evaluation of MCQ item analysis reports serves as a useful tool to detect defective MCQ items and construct a reliable MCQ bank at the department and university levels. However, not all faculty staff possess the experience and knowledge required to comprehend examination item analysis reports.¹⁻⁵ Therefore, in this study, we categorized the tested MCQ items as poor, good, and ideal, depending on the DIF I, DI, DE, and NFDs. Thus, of the 165 MCQs, 21 (12.7%) were classified as “Ideal MCQ Items” and 132 (80%) were classified as “Good MCQ Items.” Both were recommended for storage in the MCQ bank to be reused in future MCQ tests. However, of the 165 MCQ items, only 12 (7.3%) were classified as “Poor MCQ Items” and recommended for elimination from the MCQ bank or restructuring if they were to be reused in future examinations.

Similar studies have used this kind of classification. For example, *Date AP et al. (2019)*, found that of the 40 examined MCQ items, 31 items (77.5%) were excellent, with acceptable discriminating power, and they were recommended for storage in the MCQ bank for further use; only nine MCQ items (22.5%) had a poor DI and required rephrasing to be stored.¹⁹ Furthermore,, *Mehta and Mokhasi (2014)*, found that 70% of the MCQ items in their study could be classified as good

or excellent, suitable for storage in the MCQ bank.⁹ Also, in a the study done by *Mandeep Kaur et al. (2016)*, they found that 27 out of the 50 tested MCQ items were classified as being good MCQ items and recommended for storage in the MCQ bank, 17 MCQ items required revision, and only six MCQ items were recommended for elimination.¹⁰ Also the studies of *Kheyami et al. (2018)* and *Kowash et al. (2019)*, showed similar results.^{7,18} However, all of the previous studies, including the present study, agree that easy MCQ items and fair to good discrimination power with one or two NFDs and DE of 33% to 66% are suitable to be stored in the MCQ bank to be reused in future MCQ tests.

Conclusion

In our study, the majority of the tested MCQ items achieved the standards of acceptable difficulty level with fair to good discrimination power, which indicates that the chosen MCQ items for the sixth-year students studying orthodontics were of good quality. Easy MCQ items were considered good and acceptable if their DI was fair to good. MCQ items with poor discrimination or negative discrimination indices of the correct answer option were either very easy or very difficult, and were recommended for elimination or restructuring in order to create a reliable MCQ bank.

Recommendations

- Each course director should evaluate the item analysis report of each MCQ test and provide feedback on the quality of MCQ items to each faculty member involved in formulating the MCQ examination for high quality MCQ tests.
- To ensure MCQ bank excellence, all the restructured MCQ items should be compared with the item analysis results to decide if they should be reused in future tests.

Conflict of Interest

The authors declare no conflicts of interest.

References:

1. Alamoudi AA, El-Deek BS, Park YS, Al Shawwa LA, Tekian A. Evaluating the long-term impact of faculty development programs on MCQ item analysis. *Medical teacher*. 2017 Mar 16;39(sup1): S45-9.
2. Fallatah HI, Tekian A, Park YS, Al Shawa L. The validity and reliability of the sixth-year internal medical examination administered at the King Abdulaziz University Medical College. *BMC medical education*. 2015 Dec 1;15(1):10.
3. Abdel-Hameed AA, Al-Faris EA, Alorainy IA, Al-Rukban MO. The criteria and analysis of good multiple-choice questions in a health professional setting. *Saudi medical journal*. 2005 Oct;26(10):1505-10.
4. Ramakrishnan M, Sathe AB, Vinayak A. Item analysis: a tool to increase MCQ validity. *Indian J Basic Appl Med Res*. 2017; 6:67-71.
5. Namdeo SK, Sahoo B. Item analysis of multiple-choice questions from an assessment of medical students in Bhubaneswar, India. *Int J Res Med Sci*. 2016 May;4(5):1716-9.
6. Ingale AS, Giri PA, Doibale MK. Study on item and test analysis of multiple-choice questions amongst undergraduate medical students. *International Journal of Community Medicine and Public Health*. 2017 May;4(5):1562-5.
7. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple-choice questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Medical Journal*. 2018 Feb;18(1):e68.
8. Menon AR, Kannambra PN. Item analysis to identify quality multiple choice questions. *National J. Lab. y Medi*. 2017;6(2):M007-10.
9. 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.
10. Mandeep Kaur, Shweta Singla, Rajiv Mahajan. "Item analysis of in use multiple choice questions in pharmacology". *International Journal of Applied and Basic Medical Research*, Jul-Sep 2016, Vol 6, Issue 3: P 170-173
11. Rehman A, Aslam A, Hassan SH. Item analysis of multiple-choice questions. *Pakistan Oral & Dental Journal*. 2018 Sep 10;38(2):291-3.
12. 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 Jun 30;7(1):45-50.
13. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *IeJSME*. 2009;3(1):2-7.
14. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals-Academy of Medicine Singapore*. 2006 Feb 1;35(2):67.
15. Badkur M, Suryavanshi G, Abrahan AK. The Correlation between the Acceptable Range of Difficulty and Discrimination Indices in Four-Response Type Multiple Choice Questions in Physiology. *Indian Journal of*

Basic and Applied Medical Research. 2017;6(4):695-700.

16. Haladyna TM, Downing SM. A taxonomy of multiple-choice item-writing rules. *Applied measurement in education*. 1989 Jan 1;2(1):37-50.

17. Omer AA, Abdulrahim ME, Albalawi IA. Flawed multiple-choice questions put on the scale: What is their impact on students' achievement in a final undergraduate surgical examination? *Journal of Health Specialties*. 2016 Oct 1;4(4):270.

18. Kowash M, Hussein I, Al Halabi M. Evaluating the Quality of Multiple-Choice Question in Paediatric Dentistry Postgraduate Examinations. *Sultan Qaboos University Medical Journal*. 2019 May;19(2): e135.

19. Date AP, Borkar AS, Badwaik RT, Siddiqui RA, Shende TR, Dashputra AV. Item analysis as tool to validate multiple choice question bank in pharmacology. *International*

Journal of Basic & Clinical Pharmacology. 2019 Aug 28;8(9):1999-2003.

20. Gomboo A, Gombo B, Munkhgerel T, Nyamjav S, Badamdorj O. Item Analysis of Multiple-Choice Questions in Medical Licensing Examination. *Central Asian Journal of Medical Sciences*. 2019 Jun 30;5(2):141-8.

21. Rao C, Prasad HK, 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*. 2016 Oct 1;2(4):201.

22. 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 journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*. 2014 Jan;39(1):17.