DEPÓSITO LEGAL ppi 201502ZU4666
Esta publicación científica en formato digital
es continuidad de la revista impresa
ISSN 0041-8811
DEPÓSITO LEGAL pp 76-654

Revista de la Universidad del Zulia

Fundada en 1947
por el Dr. Jesús Enrique Lossada

Ciencias Exactas Naturales y de la Salud

Año 10 N° 27
Mayo - Agosto 2019
Tercera Época
Maracaibo-Venezuela

The relationship of correct option location, distractor efficiency, difficulty and discrimination indices in analysis of high-stakes multiple-choice questions exam of medical students

Madjid Shafiayan *

Balal Izanloo **

ABSTRACT

Background: Analysis of Multiple-Choice Questions (MCQs) is the psychometric **Objective:** This study was method pertains to validity and reliability of the exam. conducted to identify psychometric properties of high-stakes MCQ exam of under-graduate Medical Students' assessment. With this in mind we tried to investigate the effect of correct option location on difficulty index (DIF I) and discrimination index (DI) regarding distractor efficiency (DE) in the context of Medical Education. **Materials and Methods:** National high –stake MCQ exam was conducted to senior medical students belonging to universities of Medical Sciences to assess knowledge of Basic and Clinical sciences. Data were analyzed using Classical Test Theory to investigate effect of correct – option position on DIF I and DI and DE. Microsoft Excel spread sheet; SPSS version 23; R Psych Package softwars were used. Descriptive statistics; Point biserial correlation; Fisher's Exact Test; ANOVA test and Pearson correlation were performed. **Results:** The mean score was 107.30±19.10 ranging from 40 – 174.Mean DIF I and DI were 0.54 ± 0.20 and 0.20 ± 0.10, respectively. Fourthly three and half percent MCQs were of average DIF I (0.30< P< 0.70) and DI >0.2. Overall 127/600 (21.16%) were null distractors (< 5%) and DE was 78.84%. Mean DIF I and SD key option 1; 2; 3; 4 were 0.50±0.20; 0.59 ± 0.18; 0.54 ± 0.23; 0.50 ± 0.17, respectively. **Conclusion:** Our data suggest that correct option location remarkably affect DIF I of item. We believe our study provides considerable insight into validating MCQs of Medical students' assessment to optimizing question bank.

KEY WORDS: Difficulty index; Discrimination index; Distractor efficiency; Correct option position; Multiple – Choice Questions; Medical Education

*Ph.D Candidate Of Medical Education, Department of Medical Education,School of Medicine,Tehran University of Medical Sciences, Tehran, Iran.

**Assistant Professor of Curriculum Planning, Department of Curriculum Planning, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran.

Recibido: 11/09/2019 Aceptado: 14/10/2019

La relación de la ubicación correcta de las opciones, la eficiencia del distractor, los índices de dificultad y discriminación en el análisis de preguntas de opción múltiple de alto riesgo en el examen de estudiantes de medicina

Antecedentes: el análisis de las preguntas de opción múltiple (MCQ) es el método psicométrico relacionado con la validez y confiabilidad del examen. Objetivo: Este estudio se realizó para identificar las propiedades psicométricas del examen MCQ de alto riesgo de la evaluación de estudiantes de medicina de pregrado. Con esto en mente, tratamos de investigar el efecto de la ubicación correcta de las opciones en el índice de dificultad (DIF I) y el índice de discriminación (DI) con respecto a la eficiencia del distractor (DE) en el contexto de la educación médica. Materiales y métodos: se realizó un examen nacional MCQ de alto nivel a estudiantes de medicina de alto nivel pertenecientes a universidades de ciencias médicas para evaluar el conocimiento de las ciencias básicas y clínicas. Los datos se analizaron utilizando la teoría de prueba clásica para investigar el efecto de la posición de opción correcta en DIF I y DI y DE. Hoja de cálculo de Microsoft Excel; SPSS versión 23; Se utilizaron softwares R Psych Package. Estadísticas descriptivas; Punto de correlación biserial; Prueba exacta de Fisher; Se realizó la prueba ANOVA y la correlación de Pearson. Resultados: El puntaje promedio fue de 107.30 ± 19.10, que varió de 40 a 174.El promedio de DIF I y DI fue 0.54 ± 0.20 y 0.20 ± 0.10, respectivamente. Cuarto, el tres y medio por ciento de MCQ fue de DIF I promedio ($0.30< P< 0.70$) y DI >0.2. En general, 127/600 (21.16%) fueron distractores nulos (< 5%) y DE fue de 78.84%. Promedio DIF I y SD clave opción 1; 2; 3; 4 fueron 0,50 ± 0,20; 0,59 ± 0,18; 0,54 ± 0,23; 0.50 ± 0.17, respectivamente. Conclusión: Nuestros datos sugieren que la ubicación correcta de la opción afecta notablemente el DIF I del artículo. Creemos que nuestro estudio proporciona una visión considerable sobre la validación de MCQ de la evaluación de los estudiantes de medicina para optimizar el banco de preguntas.

PALABRAS CLAVE: índice de dificultad; Índice de discriminación; Eficiencia del distractor; Posición correcta de la opción; Preguntas de selección múltiple; Educación médica

Introduction

Assessment in Medical Education is generally accepted to be a critical component in both the teaching and learning process of Medicine. Certification procedure to pass/fail decision in high stakes exam and evaluative action are two constructs of assessment (Swanwick, 2013). MCQs are one of the assessment methods that enable assessors to measure two initial competency levels of Miller Assessment Pyramid called Knows and Knows How in clinical skills with the adoption of Bloom Taxonomy of Educational Objectives (Miller, 1990). In addition, quality assurance of the assessment method requires psychometric techniques which are applied to internal structure of test to identify flawed items in order to optimize them (Tavakol, 2016; Downing, 2003).

Item analysis typically includes computing DIF I and DI to accurately measure the examinee's abilities through developing best assessment tool. DIF I in MCQs that is dichotomously scored items is proportion of examinee who correctly responds the test item. DI that is point-biserial coefficient typically corresponds between test item response and total test score. DE reflects the number of functioning distractors (De Champlain, 2010).

This cross-sectional study aims to identify psychometric properties of high-stakes MCQ pre-internship exam of under-graduate Medical Students' assessment in Iran. Classical test theory is used to test and item analysis. With this in mind we tried to investigate the effect of key option position on DIF I and DI regarding DE in the context of Medical Education.

What we know is largely based on validating the studies that have been extensively applied to investigate psychometric properties of MCQ exam in Medical students' assessment. However, very little research to date has focused on DE relationship to DIF I and DI in high-stakes MCQ exam of Medical student's assessment. Moreover, not much is known about the association of key option position to DIF I and DI considering DE.

MCQ test and item analysis have most often been investigated in terms of psychometric properties. Soler (2002) found a schema to be useful for distractor analysis. He reported that item difficulty is related to distractors and revision of whole test and modification of distractors were needed. This left the test designer to obtain optimal approach in MCQ analysis by applying a Classical Test Theory framework (Soler and Arias, 2002).

Rodriguez (2005) furthered this idea in his Meta-analysis on empirical research and theoretical reviews by considering optimal number of multiple–choice options. Three options were found optimal for MCQ test in most settings. He also emphasized that researchers needed to study the role of more effective plausible distractors (Rodríguez, 2005).

These studies hinted that DIF I and DI may be predicted based on DE. A later experiment concluded that there was little probability of designing more than two functional distractors. As such, three-option MCQ were optimal (Tarrant et al., 2009).

A study of quality assurance of MCQ exam in high-stakes of Medical student assessment reported that the number of functional distractor in addition to DIF I and DI should be included as the criteria of MCQ exam quality (Ware, 2009).

A greater level of relationship between item difficulty and discrimination indices was found on psychometric characteristics of MCQ test in undergraduate Medical student's assessment in Malaysia (Barman et al, 2010).

Rogausch and co-workers (2010) analyzed rarely the selected distractors in high-stakes MCQ test of Medical student's assessment based on the Swiss Federal Graduation Examination since 2005 to 2007. They reported just 30% of MCQs had DE of 100%.

Another study was performed in Dental College, Karachi to find the effects of non-functioning distractors on ideal questions. They concluded that items with acceptable DIF I and DI range are those items which include at least two functional distracters (Gajjar et al, 2014).

Gajjar and co-workers (2014) analyzed MCQs of Medical student's assessment in Ahmedabad, Gujarat. They stated that MCQ item would be optimal when it has the maximum DE, high D I (DI > 0.25) and average DIF I (DIF between 31% and 60%).

Other researchers (2015) investigated the impact of low functioning distracters on MCQ test in preclinical assessment of Medical Education. Substituting the nonfunctioning distractors with functioning distractors have been demonstrated that will improve discrimination ability of the exam (Ali et al, 2015).

Another MCQ test analysis was conducted in teaching Prosthodontic to Dental students focusing on distractor analysis and item difficulty and discrimination indices. Their results highlighted that item analysis must be used to determine DIF I and DI and DE (Madhav, 2015).

Patil et al (2016) analyzed MCQ test of first year medical student assessment in India. Their results indicated that just 65% and 75% items had acceptable range of DIF I and DI.

MCQ analysis of Medical student's assessment has shown that the range of DIF I between 31% - 60%; DI (DI > 0.25) and DE of 100% are characteristics of optimal MCQ (Rao et al, 2016).

Ferdousi (2017) analyzed MCQ test of Anatomy, Biochemistry and Physiology in Bangladesh. He stressed the omission or replacement of nonfunctioning distractors.

A recent comprehensive review on MCQ analysis states that the Achilles' heels of MCQ questions has generally been DE. There have been little or no researches pertaining to key option position regarding DE (Gierl et al, 2017).

Garg and co-workers (2018) analyzed the MCQ test of Medical student assessment in India to validate DIF I and DI and DE. An association between DIF I; DI and DE was reported (Garg et al, 2018).

Kheyami and co-workers (2018) investigated the MCQ test of the Department of Pediatrics in Bahrain to assess relationship between DIF I; DI and DE. They confirmed the conclusion of previous study that the development of three-option is easier than four or five-option. Additionally, this decreased the number of nonfunctioning distractors.

A recent investigation on final summative MCQ test for the graduates of Pharmacy students in Qatar highlighted that DIF I of the exam was in an acceptable range, but DI require improvement. A high percentage of nonfunctioning distractor was obvious as a result of difficulty of faculty members to construct plausible distractors (Pawluk et al. 2018).

These studies have only dealt with psychometric properties of MCQs of Medical Student's exam, whereas our study focuses on the relationship of key option position with DIF I and DI by considering the DE of items.

The aim of the present work is a comprehensive investigation of the effect of key option location on DIF I and DI of MCQs in high-stakes pre-internship exam of undergraduate Medical Students.

The results of the present study are encouraging and have gone some way toward enhancing our understanding of the internal structure of MCQ exam in the context of Medical Education.

We hope that our research will be constructive in solving the difficulty of test construction in high-stakes MCQs Pre-Internship Exam. Furthermore, we believe that our results may improve knowledge about psychometric properties of MCQs in the field of Medical Student's assessment.

1. Method

We performed psychometric test and item analysis to investigate item and test characteristics of MCQ Pre-Internship Exam. This National high-stakes exam is carried-out to assess the competency level of Knows and Knows How category of Miller assessment Pyramid pertaining to Senior undergraduate medical students (Miller, 1990). The Exam content was blueprinted according to proportionated 17 sub-tests of clinical sciences designated to undergraduate medical education curriculum in Iran.

A total of 2758 undergraduate medical students from 43 universities of medical sciences participated in the Exam on March 6th , 2014. All examinees had to mark the correct answer out of four options on the answer sheet based on 200 MCQ randomized in two booklets in 200 minutes. There was one mark for the correct answer and no negative marking for incorrect answer.

In order not to have any direct involvement of human subjects, the examinee's identity was concealed. As such, the study was exempted from ethical oversight.

Microsoft Excel spread sheet was used to enter data that were obtained for examinees' selected response for each item from the Assessment Center of Ministry of Health and Medical Education in Iran. Data screening was executed by using SPSS version 23 software (Green et al, 2016).

In order to validate the exam, we first had to obtain descriptive statistics which include central tendency indices for example, Mean; Median; Mode, and the Range of total scores of examinees by using SPSS version 23 software (Green et al, 2016).

Our preliminary aim was to get the general picture of psychometric properties of Exam and item. The next step was consequently to investigate DIF I; DI, and DE. Classical Test Theory Model was applied using R Psych Package software to identify item parameters regarding key option position and distractors efficiency (De Champlain, 2010; Team, 2014). The percentage of the correct answer that was chosen by the examinees was considered as item difficulty. Point biserial correlation between item score and total test score was applied as item discrimination. Functional distractor option was identified if it was selected by > 5% examinees. Existence of at least 1;2, and 3 functional distractor in an MCQ item is representative of DE of 33.3%;66.6%, and 100% (De Champlain, 2010; Patil et al, 2016; Ferdousi et al, 2017; Garg, 2018; DeVellis, 2006; Team, 2014).

To provide a way of evaluating the association between DIF I and DI of exam adaptation of scatter plot was selected. This method was chosen in that it is one of the most practical ways to assess the relation between DIF I and DI of the exam. Fisher's Exact Test was used to see the associations among DIF I and DI and DE (Raymond and Rousset, 1995).

Finally, ANOVA test was used to compare the mean of DIF I, DI, and key option position. Kuder Richardson 20 was obtained for the reliability of the exam. In this study, the p values less than 0.05 were considered statistically significant (González-Rodríguez, 2012; Feldt, 1965).

2. Results

A total of 2758 undergraduate medical students fulfilled the exam. Female examinees were 1701(61.70%), whereas male students were 1057(38. 30%). Examinees' marks ranged from 40 to 174 (out of 200). The mean marks and SD were 107.30 and 19.10, respectively. Normal distribution of scores was observed. All three reliability coefficients including alpha Cronbach; alpha if item deleted, and kuder Richardson 20 were 0.89.

DIF I and DI ranged from 0.07 to 0.95 and -0.09 to 0.40, respectively. The mean DIF I and SD were 0.54 and 0.20; 95% confidence interval for the difference = 0.14 – 0.93, respectively. The mean DI and SD were 0.20 and 0.10; 95% confidence interval for the difference = 0.00 – 0.41, respectively.

Pearson correlation of the whole test, between DIF I and DI was 0.42; 95% confidence interval for the difference = 0.30 – 0.53. The correlation was significant at p ≤ 0. 001. At first, the DI rose along the DIF I. Maximum DI 0.40 occurred in acceptable DIF I range between 0.35 – 0.55 at plateau and then decreased slightly with further development in DIFI (figure 1).

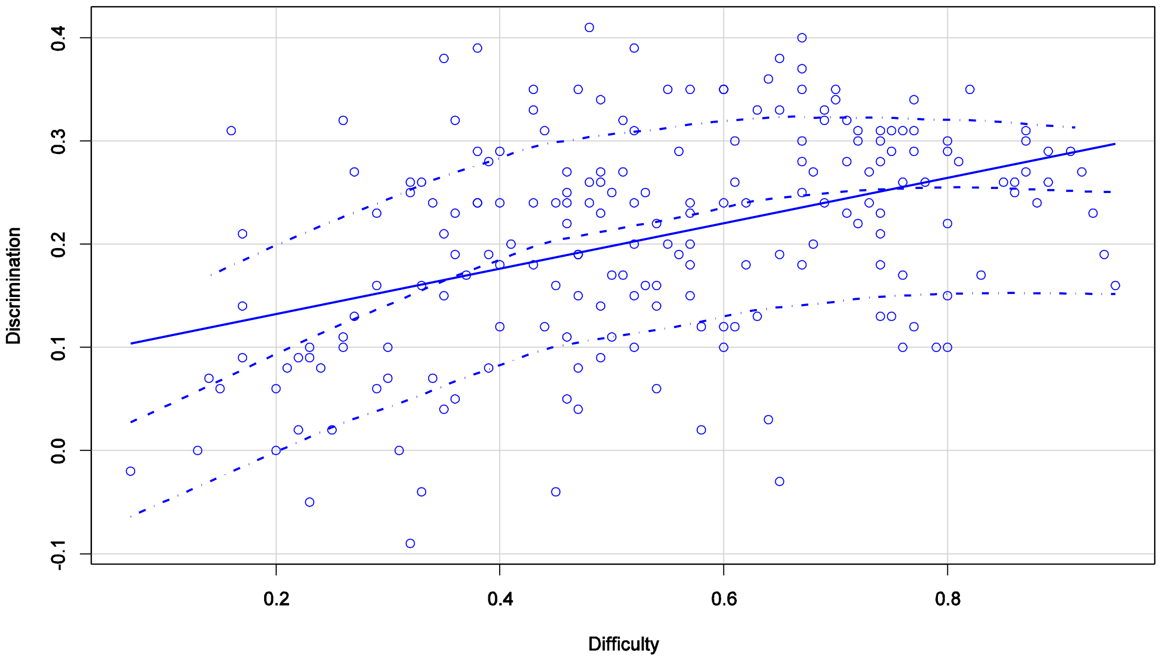

Figure 1: An approximate dome shaped regression line is not evident between DIF I and DI as illustrated on scatter plot. (N= 200)

Considering key option 1, The mean and SD for DIF I were 0.50 and 0.20, respectively. The mean and SD for DI were 0.22 and 0.09, respectively. The Fisher's Exact Test highlighted the more difficult MCQs (DI ≤ 70), the higher DE (81.81% Vs. 4.54% P = 0.01), thus association was found statistically significant. Additionally, approximately two- thirds of MCQs with key option 1 with higher DI (DIC ≥ 0.20) were having higher DE (61.36% Vs. 4.54% P = 0.31) therefore, association was found statistically insignificant (Table 1).

The mean and SD for DIF I were 0.59 and 0.18, respectively regarding key option 2. The mean and SD for DI were 0.19 and 0.10, respectively. The Fisher's Exact Test highlighted the more difficult MCQs (DI ≤ 70), the higher DE (56.85% Vs. 5.88% P = 0.01), thus association was found statistically significant. Additionally, over two-thirds of MCQs with key option 2 with higher DI (DIC ≥ 0.20) were having higher DE (35.29% Vs. 13.72% P = 0.32) therefore, association was found statistically insignificant (Table 1).

Table 1: Association among DIF I, DI and DE key 1 and key 2. (N=95)

Indices Key1 of	DE ≥ 66%	DE ≤ 33%	TOTAL	Test of Significance
DIF I ≤ 0.70	36(81.81%)	2(4.54%)	38(86.36%)	The Fisher exact test statistic value is 0.0135. The result is significant at p < .05.
DIF I > 0.70	3(6.81%)	3(6.81%)	6(13.63%)	
TOTAL	39	5	44	
DI < 0.20	12(27.27%)	3(6.81%)	15(34.09%)	The Fisher exact test statistic value is 0.3187. The result is not significant at p < .05.
DI ≥ 0.20	27(61.36%)	2(4.54%)	29(65.90%)	
TOTAL	39	5	44	
Indices Key2 of	DE ≥ 66%	DE ≤ 33%	TOTAL	
DIF I ≤ 0.70	29(56.86%)	3(5.88%)	32(62.74%)	The Fisher exact test statistic value is 0.0116. The result is significant at p < .05.
DIF I > 0.70	11(21.56%)	8(15.68%)	19(37.25%)	
TOTAL	40	11	51	
DI < 0.20	22(43.13%)	4(7.84%)	26(50.98%)	The Fisher exact test statistic value is 0.3238. The result is not significant at p < .05.
DI ≥ 0.20	18(35.29%)	7(13.72%)	25(49.01%)	
TOTAL	40	11	51	

DE = distractor efficiency; DIF I= difficulty index; DI = discrimination index

Considering key option 3, the mean and SD for DIF I were 0.54 and 0.23, respectively. The mean and SD for DI were 0.18 and 0.12, respectively. The Fisher's Exact Test highlighted almost over four-fifths of MCQs with key option 3 with lower DIF I(DIF I≤0.70) were having higher DE (65.00% Vs. 5.00% P = 0.00), thus association was found statistically significant. Additionally, half of MCQs with key option 3 with higher D I (D I ≥ 0.20) were having higher DE (40.00% Vs. 10.00% P = 1.00) therefore, association was found statistically insignificant (Table 2).

The mean DIF I and SD were 0.50 and 0.17, respectively regarding key option 4. The mean DI and SD were 0.23 and 0.09, respectively. The Fisher's Exact Test highlighted almost over four-fifths of MCQs with key option 4 with lower DIF I(DIF I≤0.70) were having higher DE (82.00% Vs. 0.00% P = 0.02), thus association was found statistically significant. Additionally, over two-third of MCQs with key option 4 with higher D I (DIC ≥ 0.20) were having higher DE (62.22% Vs. 2.22% P = 1.00) therefore, association was found statistically insignificant (Table 2).

Table 2: Association among DIF I , DI and DE key 3 and key 4. (N=105)

Indices of key 3	DE ≥ 66%	DE ≤ 33%	TOTAL	Test of Significance
DIF I ≤ 0.70	39 (65.00%)	3(5.00%)	42(70.00%)	The Fisher exact test statistic value is 0.0004. The result is significant at p < .05.
DIF I > 0.70	9(15.00%)	9(15.00%)	18(30.00%)	
TOTAL	48	12	60	
DI < 0.20	24(40.00%)	6(10.00%)	30(50.00%)	The Fisher exact test statistic value is 1. The result is not significant at p < .05.
DI ≥ 0.20	24(40.00%)	6(10.00%)	30(50.00%)	
TOTAL	48	12	60	
Indices of key 4	DE ≥ 66%	DE ≤ 33%	TOTAL	
DIF I ≤ 0.70	37 (82.22%)	0(0.00%)	37(82.22%)	The Fisher exact test statistic value is 0.0283. The result is significant at p < .05.
DIF I > 0.70	6 (13.33%)	2 (4.44%)	8(17.77%)	
TOTAL	43	2	45	
DI < 0.20	15 (33.33%)	1 (2.22%)	16(35.55%)	The Fisher exact test statistic value is 1. The result is not significant at p < .05.
DI ≥ 0.20	28 (62.22%)	1 (2.22%)	29(64.44%)	
TOTAL	43	2	45	

DE = distractor efficiency; DIF I= difficulty index; DI = discrimination index

Considering key option 1 position, Pearson correlation between DIF I and DI was 0.37; 95% confidence interval for the difference = 0.60 – 0.08. The correlation was significant at p ≤ 0. 01. Initially, DI rose along with difficulty index. Maximum DI 0.35

occurred in acceptable DIF I range between 0.40 – 0.60 at plateau and then decreased slightly with further development in DIF I (figure 2).

Pearson correlation between DIF I and DI was 0.57; 95% confidence interval for the difference = 0.73 – 0.35, regarding key option 2 position. The correlation was significant at p ≤ 0. 0001. There is an approximately linear relationship between DIF I and DI. As such, the most difficult items pertained to low discriminant ones, higher discrimination for difficulties ranged 0.30 – 0.70, and then increased slightly for the easiest items (figure 2).

Considering key option 3 position, Pearson correlation between DIF I and DI was 0.51; 95% confidence interval for the difference = 0.68 – 0.29. The correlation was significant at p ≤ 0. 0001. Discriminants were somewhat linearly associated with difficult components. Firstly, the less discriminants, the less difficult items were observed, until difficulty range between 0.70- 0.80 at plateau discrimination increased and then rose slightly for the less difficult items (figure 2).

Pearson correlation between DIF I and DI was 0.30; 95% confidence interval for the difference = 0.55 – 0.04, regarding key option 4 position. The correlation was significant at p ≤ 0. 01. First, D I increased with rise of DIF I. Maximum DI 0.35 occurred in acceptable DIF I range between 0.40 – 0.60 at plateau and then decreased slightly with further development in DIF I (figure 2).

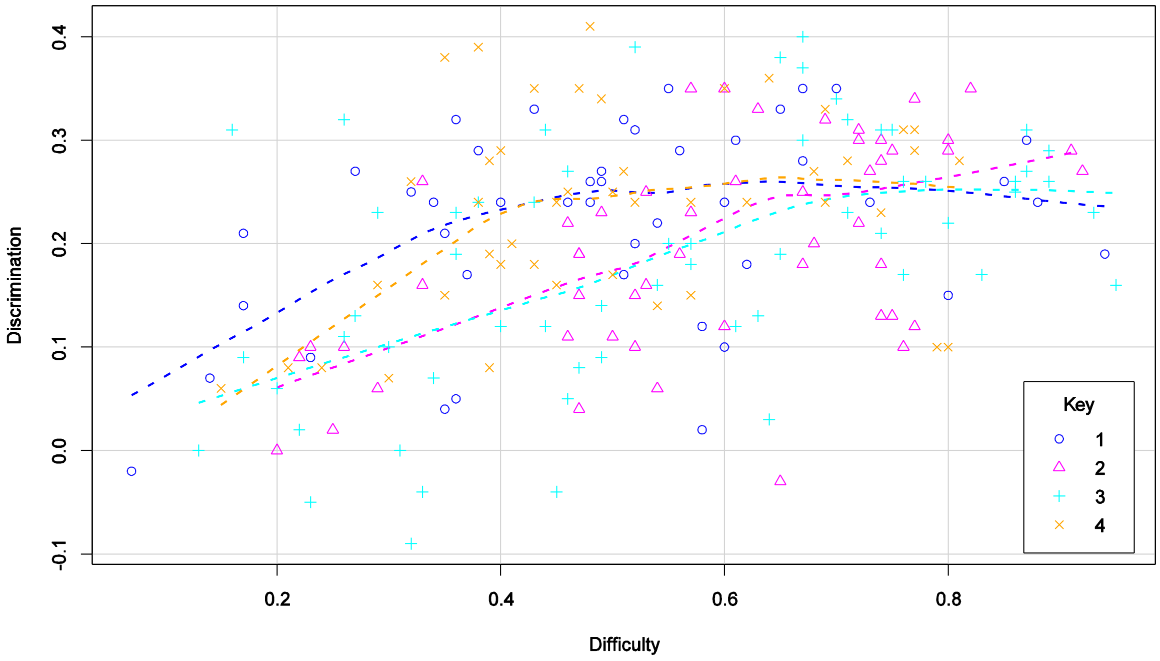

Figure 2: An approximate dome shaped regression line is not evident between difficulty and discrimination indices, regarding key option 1;2;3, and 4 as illustrated on scatter plot. (N=200)

There was not statistically significant relationship between DIF I and key option1;2;3, and 4, using ANOVA test (figure 3).

F = (3;196) = 1.857; p = 0.138

SOURCE OF VARIATION	Sum Sq	Df	Mean Sq	F value	Pr(>F)
KEY	0.223	3	0.07422	1.857	0.138
Residuals	7.832	196	0.03996		

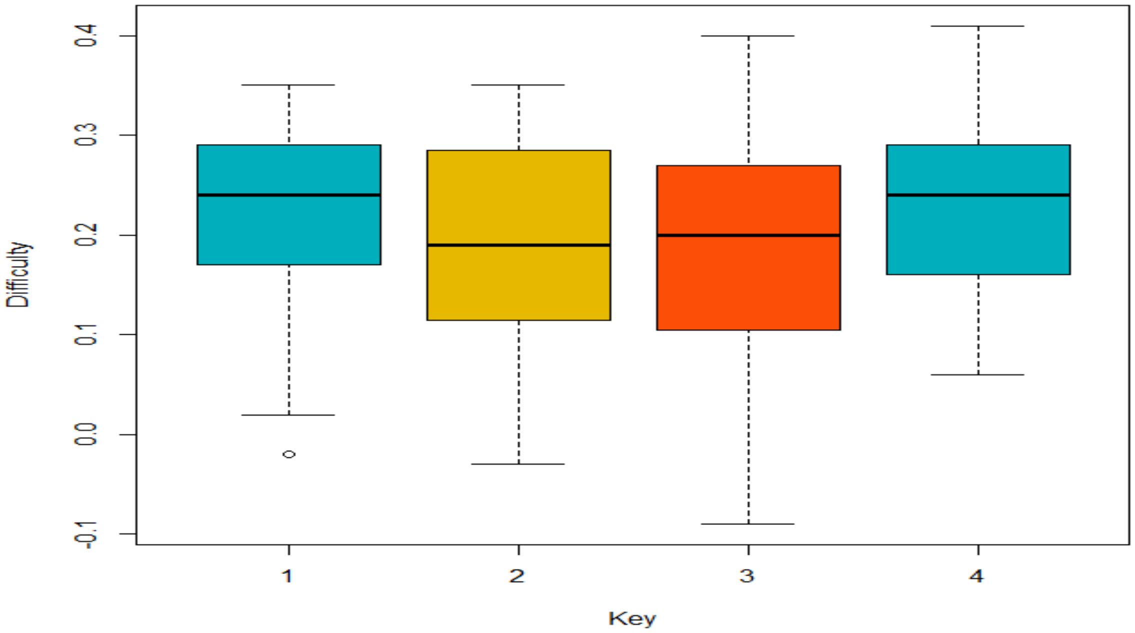

Figure 3: Relation between key option position and DIF I as reflected on Box plot illustration (n= 200) .

There was not statistically significant relationship between DI and key option1;2;3, and 4, using ANOVA test (figure 4).

F= (3;196) = 2.214; p = 0.087

SOURCE OF VARIATION	Sum Sq	Df	Mean Sq	F value	Pr(>F)
KEY	0.0721	3	0.02403	2.214	0.0878 .
Residuals	2.1267	196	0.01085		

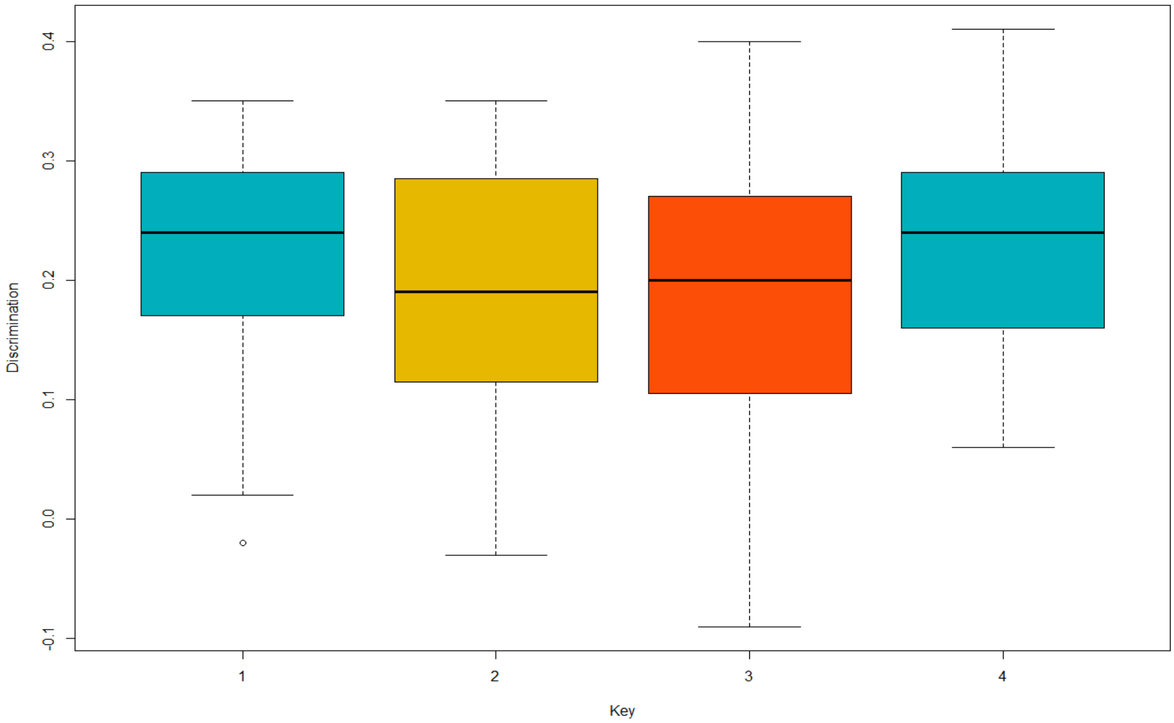

Figure 4: Relation between key option position and DI as reflected on Box plot illustration (n=200).

3. Discussion

The main goal of this study was to analyze psychometric properties of high-stakes MCQ pre-internship exam of undergraduate Medical students' assessment based on classical test theory. In addition, the present study seeks to investigate the effect of key option position on item indices regarding DE.

In our study, 123/200 items (61.5%) were within acceptable range of DIF I (0.30 – 0.70). Just 25/200 items (12.5%) were with DIF I(p<0.30). Previous studies nearly correspond with the data presented in the Results. Ferdousi (2017) reported about preclinical assessment in Bangladesh. In his analysis 43.5% and 18.67% of items had DIF I> 0.9 and 0.66-0.80, respectively (Ferdousi et al, 2017). Furthermore, similar results were obtained from another study which was conducted on summative exam of patho-physiology (Caballero et al., 2014). Gajjar (2014) found mean DIF I of 39.4 ± 21.4% in the assessment of Medical students in Pakistan. Unlike the aforementioned citation, Rao (2017) reported mean DIF I of 75.0 ± 23.7% in psychometric analysis of assessment of medical students in department of pathology. Similarly, another recent study by Garg (2019) mean DIF I of 71.6± 19.4% was reported in his analysis of formative assessment of Medical students in Delhi, India.

The mean DI and SD in our study was 0.2±0.1.However, Ferdousi (2017) concluded mean DI of 0.13. Like our research, another conducted one has reported the mean DI of case-based items was more than standard items and overall DI of 0.24 was reported (Caballero et al, 2014). In another analysis aligned with Ferdousi, Gajjar (2014) asserted mean DI of 0.14 in result of end of course exam of community medicine. Moreover, in Rao's analysis the mean DI which was more than 0.2 in 75% of items and 65% of them ≥0.4 resembles our findings. There were not any items with negative DI (Rao, et al, 2016). Likewise, mean DI and SD of 0.3±0.17 had been reported in other study by Garge (2018). According to what has been mentioned so far, the results indicate the mean DI of medical students' MCQs ranges 0.2 and 0.3.

In this study, even approximate dome shape regression line between DIF I and DI was not obvious. However, in other study by Ferdousi an illustration of dome shape regression line was obvious between DIF I and DI demonstrating the adaptability of items with maximum difficulty indices around 0.6 and DI up to 0.68. In another study, researchers reported the correlation of 60% between DIF I and DI with presence of difficult items which were not discriminating due to the very fact that difficulty level was beyond the comprehension of the class (Caballero et al, 2014). Meanwhile, Gajjar (2014) concluded in his analysis items with acceptable DIF I were not good at differentiating higher and lower ability examinees. As such, the relationship between DIF I and DI was not dome shaped. Like Ferdousi, Rao reported correlation of 0.56 between DI and DIF I indicating a reciprocal relationship. It means dome shape regression line was evident between these two indices meaning maximum discrimination power occuring in acceptable difficult items (Rao et al, 2016). Unlike the findings of the aforementioned studies, Garge in his analysis reported items with DIF I ≤ 0.7 were having higher DI, but the association were statistically insignificant (Garg, 2018). In fact, the achieved results does do not provide us with the opportunity to come up with a final decision on maximum DI occurring in acceptable DIF I range.

Ferdousi reported 15.74%; 21.31%; 12.5% as ineffective distractors in discipline of anatomy biochemistry ,physiology, respectively. Furthermore, 9.7% of distractors were with efficacy of more than 50%.(18)Gajjar reported mean DE of 88.6% ± 18.6% and 11.4% ineffective distractors.In his study , fifteen items with ineffective distractors had mean DIF I and mean DI of 53.5% ; 18% , respectively (Gajjar, 2014).(13)Rao elucidated that mean DE of 89.99% ± 24.42% with 5% ineffective distractor in his study. Thus, he observed compatibility of items including one ineffective distractor and DE 85.15% with difficulty indices (0.30 – 0.70) and discrimination indices ($D > 0.24$).(17)Garge stated mean distractor effectiveness 63.4% ± 33.3% indicating compatibility of lower difficulty indices ($p \leq 0.70$) ; higher discrimination indices ($D \geq 0.15$) with higher DE, respectively.(20)In our study, 127/600 (21.16%) were null distractors (< 5%) and distractor efficacy was 78.84%.Finally,in terms of what has been studied so far, we come up with the conclusion that in the context of Medical Student Assessment a MCQ requires at least two effective distractors to have acceptable DIF I ranging 0.3 – 0.7 and DI ≥ 0.2.

To date in the context of Medical Student Assessment, no work has been published on the role of key option position on item characteristics considering DE. Our data suggest that key option position remarkably affects DIF I of item. Furthermore, there is not noticeable relationship between key option position and discriminative index. It seems that placing key option as first and last alternative in MCQ of medical student exam can produce more slightly difficult and more discriminant items in comparing middle alternative.

The elaborated results in our analysis may have been defined as key option position characteristics in internal structure of MCQ exam, and may propose some hypotheses for research in relation to reliability and validity in this contextual Medical Student Assessment for future. We only analyzed one high–stake exam in Medical Education field. Different number of samples could lead to higher generalization of our results in the domain of assessment.

The importance of our results generality and their relative ease of application can be used in other Medical Student Assessment method, such as Objective Structured Clinical Exam (OSCE).

Reference

Swanwick TJUMEE, Theory, Practice. Understanding medical education2013. 1-6 p.

Miller GEJAm. The assessment of clinical skills/competence/performance. 1990;65(9):S63-7.

Tavakol M, Dennick RJAM. Postexamination analysis: a means of improving the exam cycle. 2016;91(9):1324.

Downing SM. Validity: on the meaningful interpretation of assessment data. Medical education. 2003;37(9):830-7.

De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. Medical education. 2010;44(1):109-17.

Soler HH, ARIAS RMJEIdlUC. A new insight into examinee behaviour in a multiple-choice test: a quantitative approach. 2002;10(2002):113-37.

Rodriguez MCJEMI, Practice. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. 2005;24(2):3-13.

Tarrant M, Ware J, Mohammed AMJBME. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. 2009;9(1):40.

Ware J, Vik T. Quality assurance of item writing: during the introduction of multiple choice questions in medicine for high stakes examinations. Medical teacher. 2009;31(3):238-43.

Barman A, Ja'afar R, Rahim F, Noor AJTomej. Psychometric Characteristics of MCQs used in Assessing Phase-II Undergraduate Medical Students of Universiti Sains Malaysia. 2010;3:1-4.

Rogausch A, Hofer R, Krebs R. Rarely selected distractors in high stakes medical multiple-choice examinations and their recognition by item authors: a simulation and survey. BMC Med Educ. 2010;10(1):85.

Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, D I and DE. JPMA The Journal of the Pakistan Medical Association. 2012;62(2):142-7.

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 Ahmedabad, Gujarat. Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine. 2014;39(1):17.

Ali SH, Ruit KG. The Impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. Perspectives on medical education. 2015;4(5):244-51.

Madhav V. Item Analysis of Multiple-Choice Questions in Teaching Prosthodontics. Journal of dental education. 2015;79(11):1314-9.

Patil PS, Dhobale MR, Mudiraj NR. ITEM ANALYSIS OF MCQS'-MYTHS AND REALITIES WHEN APPLYING THEM AS AN ASSESSMENT TOOL FOR MEDICAL STUDENTS. International Journal of Current Research and Review. 2016;8(13):12.

Rao C, Kishan Prasad H, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. 2016;2(4):201-4.

Ferdousi S, Rahman MM, Talukder HK, Habib MA. Post-application Quality Analysis of MCQs of Preclinical Examination Using Item Analysis. Bangladesh Journal of Medical Education. 2017;7(1):2-7.

Gierl MJ, Bulut O, Guo Q, Zhang XJRoER. Developing, analyzing, and using distractors for multiple-choice tests in education: a comprehensive review. 2017;87(6):1082-116.

Garg R, Kumar V, Maria J. Analysis of multiple choice questions from a formative assessment of medical students of a medical college in Delhi, India. International Journal of Research in Medical Sciences. 2018;7(1):4.

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;18(1):e68.

Pawluk SA, Shah K, Minhas R, Rainkie D, Wilby KJ. A psychometric analysis of a newly developed summative, multiple choice question assessment adapted from Canada to a Middle Eastern context. Currents in Pharmacy Teaching and Learning. 2018.

Veney JE. Statistics for health policy and administration using Microsoft Excel: Jossey-Bass San Francisco; 2003.

Green SB, Salkind NJ. Using SPSS for Windows and Macintosh, Books a la Carte: Pearson; 2016.

Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. 2014.

DeVellis RFJMc. Classical test theory. 2006:S50-S9.

Raymond M, Rousset FJE. An exact test for population differentiation. 1995;49(6):1280-3.

González-Rodríguez G, Colubi A, Gil MÁJCS, Analysis D. Fuzzy data treated as functional data: A one-way ANOVA test approach. 2012;56(4):943-55.

Feldt LSJP. The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. 1965;30(3):357-70.

Caballero J, Wolowich WR, Benavides S, Marino J. Difficulty and discrimination indices of multiple-choice examination items in a college of pharmacy therapeutics and pathophysiology course sequence. Int J Pharm Pract. 2014;22(1):76-83.