Educational Research & Innovation Journal
Issued by Education Strategy Administration
Ain Shams University

مجلة بحوث التعليم والابتكار
تصدر عن ادارة تطوير التعليم
جامعة عين شمس

Evaluating the effectiveness of the response - after MCQ exam item analysis on students' performance in a Communication Skills exam

تقييم فاعلية الاستجابة – بعد تحليل عناصر اختبارات الاختيار من متعدد– على أداء الطلاب في اختبار مهارات الاتصال

Rania Moustafa Hadhoud[1]
[1]Lecturer - Forensic and toxicology department - Faculty of medicine - Ain shams university

Abstract

Background: Designing an adequately structured MCQ is both complicated and time-consuming. Even after creating appropriately constructed MCQs, item analysis still needs to verify them. **Objectives:** The purpose of this study was to assess the progression of students' results across three different consecutive exams following the author's constructed response, using post-exam MCQs item analysis. Addressing potential exam flaws leads to appropriate action to improve the test standard quality, such as keeping and editing the questions or discarding them. **Materials and Methods:** This cross-sectional descriptive study was conducted on a total of 138 students subjected to three different rotational exams in the Communication Skills course during the academic year 2019-2020. Items were analyzed according to their difficulty index (DIFI), discrimination index (DI), and point biserial correlation values (PBS). Accordingly, action was taken to improve items construction, and students' results were assessed. **Results:** The three exams showed a gradual increase in the reliability scores. The mean scores showed a significant progression over time. The distribution of DIFI differed significantly among the three exams. Better values of PBS and DI were demonstrated in exams B and C compared to exam A. **Conclusion:** Constructed response according to item analysis is of considerable value in promoting the MCQs standard quality. This is obtained by affirming acceptable DIFI, PBS, and DI indices. The item analysis conducted in this study revealed that communication skills exams showed a considerable rate of acceptable DIFI, PBS, and DI values, particularly in the later tests.

Keywords: MCQs, reliability, item analysis, difficulty index, discrimination index, point biserial correlation.

المستخلص:

الخلفية: يعد تصميم أسئلة الاختيار من متعدد المهيكلة بشكل مناسب أمرًا معقدًا ويستغرق وقتًا طويلاً. حتى بعد إنشائها بشكل مناسب، لا يزال تحليل العناصر بحاجة إلى التحقق منه. الأهداف: هدفت هذه الدراسة إلى تقييم التقدم المحرز في نتائج الطلاب عبر ثلاث اختبارات متتالية مختلفة، بعد استجابة الباحث، وفقًا لتحليل عناصر ما بعد الاختبار. تؤدي معالجة العيوب المحتملة في الامتحان إلى اتخاذ إجراء مناسب لتحسين جودة الاختبار القياسية، إما الاحتفاظ بالأسئلة وتحريرها أو التخلص منها. المواد والطرق: أجريت هذه الدراسة الوصفية المقطعية على إجمالي 138 طالبًا خضعوا لثلاثة اختبارات مختلفة في مقرر مهارات الاتصال خلال العام الدراسي 2019–2020. تم تحليل العناصر وفقًا لمؤشر الصعوبة، ومؤشر التمييز، وقيم الارتباط ثنائي التسلسل. وفقًا لذلك، تم اتخاذ إجراءات لتحسين بناء العناصر، وتم تقييم نتائج الطلاب. النتائج: أظهرت الاختبارات الثلاثة زيادة تدريجية في درجات الموثوقية. تم العثور على تقدم تدريجي كبير في متوسط الدرجات. تم العثور على فرق كبير بين الاختبارات الثلاثة في توزيع مؤشر الصعوبة. تم إظهار قيم أفضل لقيم الارتباط ثنائي التسلسل ومؤشر التمييز في الاختبارين B وC مقارنة بالامتحان A. الخلاصة: تعتبر الإجابة المركبة وفقًا لتحليل العناصر ذات قيمة كبيرة في تعزيز الجودة القياسية لأسئلة الاختيار من متعدد. يتم الحصول على ذلك من خلال تأكيد مؤشرات الصعوبة وقيم الارتباط ثنائي التسلسل ومؤشر التمييز المقبولة. أظهر تحليل العنصر الذي تم إجراؤه في هذه الدراسة أن اختبارات مهارات الاتصال أظهرت معدلًا كبيرًا لهذه القيم المقبولة، لا سيما في الاختبارات اللاحقة.

الكلمات المفتاحية: أسئلة الاختيار من متعدد – الموثوقية – تحليل العناصر– مؤشر الصعوبة – مؤشر التمييز – الارتباط ثنائي التسلسل النقطي.

INTRODUCTION

The most common objective evaluation form for medical students is multiple-choice questions (MCQs). MCQs are tests that consist of multiple-choice questions followed by multiple-answer options, with only one answer option serving as the key answer and the others serving as distractors.[1] The correct or best response is the key answer, while the distractors are essentially incorrect responses.[2]

MCQs assessment has the advantage of allowing more extensive sampling, with the evaluation of the capabilities of a large sample of the students in various aspects and comprehensive subject coverage during a relatively short time.[3]

Furthermore, this assessment can reveal the regions of difficulties encountered by the students when the design of the incorrect responses demonstrates the common mis-concepts faced by the students. Another benefit of the MCQ assessment is that it is easily applicable and adaptable to the computer delivery system. MCQs assessment results are more reliable than essay tests due to their ease of discrimination and scoring of performance levels.[4]

However, creating adequately structured MCQs that assess students' interpretation and application of acquired knowledge is a complicated, difficult, and time-consuming process.[5]

Even after creating such appropriately constructed MCQs assessments, they must be verified for compliance with quality standards and affirmed coverage of all aspects of the relevant subjects.[6]

An item analysis is a process that assesses the validity and reliability of an MCQ test. This is accomplished by evaluating the student's performance on each MCQ item and statistically analyzing the likelihood of keeping, reviewing, or discarding the item from the test. [7]

For item analysis, several parameters are available, the most common of which are the difficulty index (DIFI), point biserial correlation (PBS), and discrimination index (DI). DIFI displays the percentage of correct answers as a percentage of total responses.[8] DI distinguishes between different students' achievement levels; it ranges between -1.00 and +1.00.[9] PBS correlates the students' correct/incorrect responses to each item with the total scores of the remaining items.

The PBS values range from -1.0 to +1.0 [10]. An ideal item should have a DIFI value of 30 to 70%, a DI value greater than 0.2[8], and a PBS value of at least 0.25.[10]

To achieve learning validity, the educational process goals must be aligned with the test construction. Previous research looked into the significance of continuous assessment of MCQs using item analysis. It was discovered that this could be a method for improving test construction, resulting in more valid teaching methods.[4] Item analysis was put to the test in a variety of subjects and courses. There was, however, no item analysis in the field of student development and communication skills.

The purpose of this study was to assess the progress of students' results across three different consecutive exams, after constructed response, using post-exam MCQs item analysis. Addressing potential exam flaws leads to appropriate action to improve the test standard quality, such as keeping and editing the questions or discarding them.

MATERIALS AND METHODS

This is a cross-sectional descriptive study that was conducted in the Student Development Unit, the Medical Education Development Unit, Armed Forces Collage of Medicine (AFCM), Cairo, Egypt, after the approval of the research ethics committee and Head of Medical Education Department- AFCM

Second-year medical students, Bachelor of Medicine, Bachelor of Surgery (MBBS), were targeted. The cadets had a course of

Communication Skills as a part of the Non-Technical Skills Module integrated into the internship curriculum in AFCM during the academic year 2019-2020. Each group of 46 students had their course on four consecutive days at different timings throughout the academic year. For each group of students, the same educational methods were followed by the same instructors with a total of 138 students. Then, each group was subjected to a different exam. Each exam contained 40 items of the single best response type. A blueprint of the course exams was first developed, and the subject experts prepared the exams. Pre-validation of the exams was done by faculty experts (AFCM exam committee). Each of the MCQ items had one key answer and three distractors. Each correct response had a score of 1. The time allowed for each exam was 45 minutes.

Item analysis

Item Post-validation and psychometric test analysis were performed for each exam at the Medical Education Department, AFCM. Each exam was corrected, and grades were tabulated. Afterward, the scores obtained by the students were arranged in descending order in Microsoft office excel sheet 2016. For item analysis purposes, the exam takers were categorized into three groups according to their scores on the test as a whole; an upper group consisting of 12 students (27%) who made the highest scores, a lower group consisting of 12 students (27%) who made the lowest scores and a middle group comprised of the remaining 22 students (46%). Afterward, we included the upper and lower groups and excluded the middle ones for the calculated parameters. Items were then categorized according to their difficulty index, discrimination index, and point biserial correlation values. Mean, standard deviation, and the correlation coefficient were obtained. The defective items were examined further to ensure proper construction and optimization.

Calculated Parameters

- ***Test Score Reliability:*** The Exam Analysis Report uses the Cronbach's Alpha measure of internal consistency for multiple-choice items, which provides reliable information about items scored dichotomously (i.e., correct/incorrect) (Table 1) .[11]

Table 1: Cronbach's Alpha interpretation of test score reliability in multiple-choice items

Test Score Reliability	Interpretation
< 0.50	Questionable
0.50-0.60	Revise the test
0.60-0.70	Low
0.70-0.80	Good
0.80-0.90	Very good
>0.90	Excellent

- ***Difficulty Index:*** The formula for the item difficulty index is p =c/n x 100, where c is the number of students who selected the correct answer and n is the total number of respondents. It ranges from 0% to 100% or maybe written as a proportion 0.00 to 1.00 (Table 2).[8]

Table 2: Difficulty index interpretation

Difficulty index	Interpretation	Recommendations
< 20	Very difficult	Should be revised
20-90	Good	Retained in the Questions bank
>90	Very easy	carefully reviewed

- ***Point-Biserial correlation (PBS):*** is calculated via statistical correlation of each item score to the summation of the remaining item scores. It ranges from -1.00 to 1.00 (Table 3). [10]

Table 3: PBS interpretation

PBS value	Interpretation	Recommendations
Negative	Defective items/wrong key	Discard
<0.1- 0.25	Poor	Revision for incorrect key/ discard
0.25-0.40	Good and acceptable	Scope of improvement
> 0.40	Excellent	Retained in the Questions bank

- ***Discrimination index (DI)*** was estimated using the following formula: D = PU-PL, where PU and PL are the proportions of the upper and lower group students who got the item correct. It ranges between -1.00 to 1.00 (Table 4).[9]

Table 4: Discrimination index interpretation

DI	Interpretation	Recommendations
Negative	Defective items/wrong key	Discard /correct the key
< 0.19	Poor discrimination	Discard
0.20-0.29	Marginal items	Reviewed
0.30-0.39	reasonably good but	Scope of improvement
> 0.4	Excellent discrimination	Retained in the Questions bank

Outcomes of the study

The primary outcome of this study was the differences in the indices used for item analysis among the three different exams, and the secondary outcome was the potential progress of students' results across the three sequential MCQs exams after performing recommendations based on item analysis results. Data were statistically analyzed using the software Statistical Package for the Social Sciences, version 22 (IBM, Armonk, New York, United States). A P-value less than 0.05 was considered significant.

RESULTS

Tests scores and reliability

The three exams' reliabilities and their interpretation are presented in table 5; a gradual increase in the reliability scores across the three exams was noted.

Table 5: Reliability scores of the three exams.

Exam	Reliability score	Interpretation
Exam A	0.52	Reliability is pretty acceptable, most probably due to many flawed questions as regard s design or teaching misperception
Exam B	0.64	Reliability is relatively acceptable but more than the previous exam, most probably due to teaching misperception
Exam C	0.75	Reliability could be acceptable but more than the previous exam.

The Exams Item analysis and the Results of the Students in the three exams

Item analysis of the three exams revealed that, in terms of DIFI, exam B and C had no very difficult items; all of the items were good and very easy, with a significant difference in the distribution of DIFI among the three exams. Regarding the PBS, better values were demonstrated in exams B and C compared to exam A. However, the difference is statistically non-significant. As for the DI, a higher frequency of excellent discrimination and a lower frequency of poor discrimination were shown in exams B and C compared to exam A; again, the difference did not reach the significance level (Table 6).

The students' results of the three exams are presented in table 6. No one failed in either of the exams. No significant difference was demonstrated among the three exams in distributing the result grades: excellent, very good, good, poor, or fail.

Table 6: Students results and item analysis of the three exams

	Exam A N (%)	Exam B N (%)	Exam C N (%)	X^2	p-value
Difficulty Index (DIFI)					
< 20 (Very difficult)	5 (12.5)	0 (0)	0 (0)	10.74	0.03*
20-90 (Good)	26 (65)	30 (75)	28 (70)		
>90 (Very easy)	9 (22.5)	10 (25)	12 (30)		
Point Biserial Correlation (PBS) value					
Negative (Defective items/wrong key)	2 (5)	1 (2.5)	1 (2.5)	9.11	0.17
<0.1- 0.25 (Poor)	21 (52.5)	13 (32.5)	10 (25)		
0.25-0.40 (Good and acceptable)	11 (27.5)	16 (40)	15 (37.5)		
> 0.40 (Excellent)	6 (15)	10 (25)	14 (35)		
Discrimination Index (DI)					
Negative (Defective items/wrong key)	1 (2.5)	0 (0)	1 (2.5)	3.7	0.88
< 0.19 (Poor discrimination)	20 (50)	15 (37.5)	14 (35)		
0.20-0.29 (Marginal items)	5 (12.5)	7 (17.5)	7 (17.5)		
0.30-0.39 (reasonably good)	7 (17.5)	8 (20)	7 (17.5)		
>0.4 (Excellent discrimination)	7 (17.5)	10 (25)	11 (27.5)		
Results of the students					
Excellent % (85-100)	34	37	30	3.26	0.78
Very good% (75-84.99)	31	35	41		
Good% (65-74.99)	20	17	16		
Poor % (40-64.99)	15	11	13		
Fail (0-39.99)	0	0	0		

X^2: Chi-square test, *: statistically significant.

The mean and SD of the scores in the three exams are demonstrated in table 7. Gradual progression in the mean scores was found. A significant difference was noted among the three exams. When conducting a posthoc test, it was pointed out that there were significant differences between mean scores of A & B and those of A & C, while no significant differences were noted between mean scores of B & C.

Table 7: Mean scores of the three tests

	Exam A	Exam B	Exam C	F	p-value
Mean	27.33	30.98	31.54	15.62	0.00**
SD	3.24	3.94	4.49		
Tukey Post-hoc test	A vs. B (P=0.0001)	B vs. C (P=0.77)	C vs. A (P=0.00)		

F: ANOVA test, **: statistically highly significant.

Notes about exam A

It was noted that none of the examinees selected the key answer in item 3; after revision, it was found that the key was faulty, and the item was remarked.

The wrong distractors in items 12 and 39 were chosen by high performers more than low performers, while the key answers were chosen by low performers more than high performers; this was demonstrated by negative item statistics (DI= - 0.08 and PBS= - 0.06 for item 12 and DI =0 for item 39). This raised the possibility that the students misinterpreted the questions as a result of a flaw in their design. As a result, the questions were removed from the final score and revised for improvement.

Item 28 and 29 were extremely difficult because only 9% of the examinees chose the correct answers in both items. Item 28 had an acceptable PBS, whereas item 29 had a discrimination index of zero and a PBS of 0.02. This was regarded as a teaching blunder. During the lectures, more clarification of the topics concerning both items was required. Item 28 was redesigned to be applicable for another examination and was retained in the score, whereas item 29 was removed from the score and revised for improvement.

Notes about exam B

Although item 30 was answered correctly by 65% of the students, it did not discriminate between high and low performers (DI=0/PBS=0.4). There was a flaw in the design of the item. Accordingly, the question was revised for improvement.

Notes about exam C

Although item 19 was answered correctly by 89% of the students, it did not discriminate between high and low performers (DI=-0.08/PBS= -0.02). There was a flaw in the design of the item. Accordingly, the question was revised for improvement.

Following the item analysis of exam, A, the author implemented the recommendations in tables 1,2,3, and 4 for improving question construction in the subsequent exams (B and C), such as selecting plausible distractors, informing course instructors about any teaching defects or student misinterpretation, and removing flawed questions from the questions bank.

DISCUSSION

Item-writing flaws are flaws that can occur in MCQ assessments (IWFs). Although these IWFs appear to be minor, they can significantly impair students' ability to understand and respond to questions. [12] In this way, IWFs can have an effect on the reliability of total test scores. [13]

As a result, it is advised to support efforts aimed at improving the quality of MCQs. Medical students are especially concerned because consistent preparation is ethically obligatory. Students who pass an improperly constructed exam, despite lacking adequate knowledge of the subject matter, are ultimately a hampered threat to society.

These considerations were the motivation for the current work, which aimed to evaluate the progress of students' results across three consecutive exams, with post-exam MCQs item analysis to address potential exam flaws that would help to take the appropriate action to improve the test standard quality, either by keeping the questions, editing them, or discarding them.

The current study revealed a gradual increase in the reliability scores across the three exams. This reflects the continuous steps taken after each exam to improve its quality. These results align with the study of Ali et al. who stated that administering two versions of an exam offers trainees and instructors the chance to depict areas that need improvement .[14]

In the current study, item analysis of the three exams revealed that the DIFI had the highest percentage of items (70%) in the range of 20-90 percent, indicating good (acceptable) difficulty, while the others were very difficult (4.2 percent) and very easy (4.2 percent) (25.8 percent). Exams B and C had no very difficult items; all of the items were good and very easy, with a significant difference in the distribution

of DIFI between the three exams. These findings seem to mirror the continuous improvement of MCQs construction by the AFCM exams committee.

Lin et al. reported a DIFI range of 10–93 percent, which is consistent with our findings.[15]
Karelia et al. discovered that 61, 24, and 15% of pharmacology exam items were good, very easy, and very difficult, respectively.[16] In Kheyami et alstudy, .'s these grades of difficulty were shown in 53.4, 25.9, and 20.8 percent of the items, respectively, and revealed improvement in the DIFI across consecutive exams in Pediatrics, which is consistent with our findings; they attributed this to the examination committee's efforts in the MCQs construction. [17]

Researchers prefer creating MCQs with lower DIFI values for the main topics that students may be familiar with. In addition, commencing the exam with such questions will elevate the students' confidence. Within the same context, high DIFI MCQs are better to be near the end of the exam to help discriminate between high and low achievers .[6]

Regarding the PBS, better values were demonstrated in exams B and C compared to exam A. However, the difference is statistically non-significant. In terms of DI, exams B and C showed a higher frequency of excellent discrimination and a lower frequency of poor discrimination when compared to exam A; however, the difference did not reach the significance level.

Both PBS and DI reflect the discrimination power of an item. The current study found that in the first exam (exam A), 60 and 57.5 percent of the total items had acceptable PBS and DI scores, respectively; these percentages increased to 65 and 62.5 percent in exam B and 72.5 and 62.5 percent in exam C. When considering PBS and DI values, negatively scored items accounted for 3.3 and 1.7 percent, respectively.

Our figures are comparable with those reported by Rao et al. [7.] They found that 75% of the items in the Pathology exam showed a discrimination index higher than 0.2. In variance with our results, no items were reported with a negative discrimination index in their study. Also, Musa et al. reported a percentage of negatively scored items of 0.8% in a Physiology exam.[18]

However, other previous studies have found more negatively scored items than we did, with Gajjar et al. claiming 20%.[9] Our figures are also lower than those reported in previous studies, which ranged from 4-23 percent. [19-23]

Items with negative discrimination indices reduce exam validity and should be removed. This was done as part of our preparation for the Communication Skills exams.

The students' results of the three different exams showed that no one failed in either of the exams. No significant difference was demonstrated among the three exams in distributing the result grades: excellent, very good, good, poor, or fail. Nonetheless, a higher percentage of students achieved very good and excellent grades collectively in exams B and C (72 and 71%, respectively) compared to exam A (65%). This was ensured by the significantly improved reliabilities of Exams B and C compared to exam A found in this study.

In harmony with our findings, Barry et al. reported that continuous targeted review sessions that are dedicatedly designed are a valuable tool in improving the students' achievement across consecutive exams .[24.]

Further validation tests on other modules constructed by the Medical Education Development Unit are recommended. They will be more reproducible if conducted on a larger sample of students. Other item analysis indices such as distractors efficiency analysis should be considered in further studies.

CONCLUSION

Item analysis is of considerable value in promoting the MCQs standard quality. This is obtained by affirming acceptable DIFI, PBS, and DI indices. The item analysis conducted in this study revealed that communication skills exams showed a considerable rate of acceptable DIFI, PBS, and DI values, particularly in the later tests. However, further modulation is needed to improve construction and impede the negatively scored items.

Acknowledgments

The author would like to thank Prof. Dr. Omaima Hamed, Dean of Medical Education Department, Armed Forces College of Medicine (AFCM), for all the support she provided.

Disclaimer

The article has not been previously presented or published and is not part of a thesis project.

Conflict of Interest

There are no financial, personal, or professional conflicts of interest to declare.

REFERENCES

1. Przymuszała P, Piotrowska K, Lipski D, Marciniak R, Cerbin-Koczorowska M. Guidelines on Writing Multiple Choice Questions: A Well-Received and Effective Faculty Development Intervention. SAGE Open. 2020 Jul 1;10(3). https://doi.org/10.1177%2F2158244020947432
2. Shin J, Guo Q, Gierl MJ. Multiple-Choice Item Distractor Development Using Topic Modeling Approaches. Frontiers in Psychology. 2019; 10:825. https://doi.org/10.3389/fpsyg.2019.00825
3. Fozzard N, Pearson A, du Toit E, Naug H, Wen W, Peak IR. 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 Nov 7;18(1):252. https://doi.org/10.1186/s12909-018-1346-4
4. Gupta C, Jain A, D'Souza AS. Essay versus multiple-choice: A perspective from the undergraduate student point of view with its implications for examination. Gazi Medical Journal. 2016;27(1):8–10. doi:http://dx.doi.org/10.12996/gmj.2016.03
5. Elgadal AH, Mariod AA. Item Analysis of Multiple-choice Questions (MCQs): Assessment Tool for Quality Assurance Measures. Sudan Journal of Medical Sciences (SJMS). 2021 Sep 30;334–46. doi: 10.18502/sjms.v16i3.9695
6. Scully D. Constructing Multiple-Choice Items to Measure Higher-Order Thinking. Practical Assessment, Research, and Evaluation. 2019 Nov 25;22(4): 1-13. https://doi.org/10.7275/swgt-rj52
7. Rao C, Kishan Prasad HL, 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 research. 2016;2(4):201-4. doi: 10.4103/2395-2296.189670
8. Ali I, Haral MN, Tahira F, Ali M, Imran A. An Evaluation, Assortment, and Optimization of the Question Items. Education and Urban Society. 2020 Nov 1;52(8):1264–80. https://doi.org/10.1177%2F0013124519896831
9. 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 J Community Med. 2014;39(1):17–20. doi: 10.4103/0970-0218.126347
10. Kornbrot D. Point Biserial Correlation. In: Wiley StatsRef: Statistics Reference Online [Internet]. John Wiley & Sons, Ltd; 2014 [cited 2021 Nov 15]. Available from. https://doi.org/10.1002/9781118445112.stat06227
11. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ. 2018 Dec 1;48(6):1273–96. doi: 10.1007/s11165-016-9602-2
12. Tariq S, Tariq S, Maqsood S, Jawed S, Baig M. Evaluation of Cognitive levels and Item writing flaws in Medical Pharmacology Internal Assessment Examinations. Pak J Med Sci. 2017 Jul-Aug;33(4):866-870.doi: 10.12669/pjms.334.12887
13. Nedeau-Cayo R, Laughlin D, Rus L, Hall J. Assessment of item-writing flaws in multiple-choice questions. J Nurses Prof Dev. 2013 Mar-Apr;29(2):52-7; quiz E1-2. doi: 10.1097/NND.0b013e318286c2f1

14. Ali SH, Carr PA, Ruit KG. Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. Journal of the Scholarship of Teaching and Learning. 2016 Feb 26;16(1):1–1 doi: 10.14434/josotl.v16i1.19106
15. Lin LC, Tseng HM, Wu SC. Item analysis of the registered nurse license exam by nursing candidates from vocational nursing high schools in Taiwan. Proc Natl Sci Counc Repub China D. 1999; 9:24–31. doi: 10.18295/squmj.2018.18.01.011
16. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of year II M.B.B.S students. Int EJ Sci Med Educ 2013; 7:41–6.
17. 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 Univ Med J. 2018 Feb;18(1):e68-e74. doi: 10.18295/squmj.2018.18.01.011
18. Musa A, Shaheen S, Elmardi A, Ahmed A. Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. Khartoum Medical Journal. 2018; 11(2):1477–86.
19. Shete A, Kausar A, Lakhkar K, Khan S. Item analysis: An evaluation of multiple-choice questions in Physiology examination. Journal of Contemporary Medical Education. 2015;3:106-9. doi:10.5455/JCME.20151011041414
20. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. South-East Asian Journal of Medical Education.2013; 7:45-50. doi: 10.9790/7388-0705042843
21. Elfaki OA, Bahamdan KA, Al-Humayed S. Evaluating the quality of multiple-choice questions used for final exams at the Department of Internal Medicine, College of Medicine, King Khalid University. Sudan Medical Monitor. 2015;10:123-7. doi: 10.4103/1858-5000.171863
22. Halikar SS, Godbole V, Chaudhari S. Item Analysis to Assess Quality of MCQs. Indian Journal of Applied Research. 2016;6:123-5. doi:10.36106/ijar
23. Patil R, Palve SB, Vell K, Boratne AV. Evaluation of multiple-choice questions by item analysis in a medical college at Pondicherry, India. International Journal of Community Medicine and Public Health. 2016;3:1612-6. doi:10.18203/2394-6040.IJCMPH20161638
24. Barry O, O'Sullivan E, McCarthy M. Periodic review sessions contribute to student learning across the disciplines in pharmacology. Journal of the Scholarship of Teaching and Learning. 2014 Feb 2;15:38. doi: 10.14434/josotl.v15i1.