Scholarship of Teaching Open Access

Investigating the Impact of a Multiple Choice Questions Standardization Workshop on the Performance and Satisfaction of Nursing and Midwifery Faculty Members for Designing the Final Examination Questions and Student Satisfaction in Zanjan University of Medical Sciences

Azam Maleki [1], Zeinab Ghahremani [*2], Fatemeh Ghorbani [3]

[1] Social Determinants of Health Research Center, Zanjan University of Medical Sciences, Zanjan, Iran.
[2] Psychiatric Nursing Group, Zanjan University of Medical Sciences, Zanjan, Iran.
[3] Medical- surgical Nursing Group, Zanjan University of Medical Sciences, Zanjan, Iran.

Article Info

Article history:
Received 8 Oct 2019
Accepted 8 Jun 2020
Published 18 Nov 2020
Keywords:
Standardization of Multiple-choice Questions
Designing Multiple-choice Questions

***Corresponding author:**
Zeinab Ghahremani, Psychiatric Nursing Group, Zanjan University of Medical Sciences, Zanjan, Iran.
Email: ghahremani@zums.ac.ir

Abstract

Background & Objective: Designing standard Multiple-Choice Questions (MCQs) can enhance education quality. This study aimed to determine the effect of holding a workshop on the standardization of MCQs for the final examination of the performance and satisfaction of faculties and student satisfaction in the Zanjan University of Medical Sciences.

Materials and Methods: This was a pre-post single-group interventional study performed during 2018-2019. The performance of the faculties was assessed in terms of observing the structural rules and taxonomic level of the questions in the two stages before the consecutive semesters after the intervention. In addition, the level of satisfaction was evaluated using a researcher-made questionnaire. The sample population consisted of 50 faculty members selected by a simple random method. The educational content was composed of the objective questions design principles, face and content validity, and difficulty/discrimination indices of the questions. Moreover, data analysis was performed applying descriptive statistics and chi-square at 95% confidence interval.

Results: This study reviewed 932 and 1679 questions before and after the intervention, respectively. There was a higher satisfaction level for all the Millman checklist criteria after intervening in the structure of questions.

As a result, the total mean score of the structure of the questions increased from 12.67±0.86 to 13.0±06.69, demonstrating a statistically significant increase in this regard ($P \leq 0.001$). The comprehension and knowledge questions had the highest, and evaluation questions had the lowest percentage per Bloom's taxonomy in each semester. The satisfaction level of the tutors increased more than 50% after the workshop, and more than 50% of students were satisfied with the quality of the final examination questions.

Conclusion: Given the improved performance and satisfaction of the teachers after the intervention, it is suggested to run empowerment workshops continuously.

Introduction

The Multiple-Choice Questions (MCQs) are a significant part of the higher education exam questions. Therefore, it is advised to observe the content validity of knowledge level (taxonomy) and fundamental rules of the question in designing final examinations so it will be possible to distinguish between weak and strong students. Otherwise, it will reduce learners' motivation and waste the efforts of tutors and the education system (1, 2). Besides, the type and quality of the final examination are affective on the teacher's teaching method and credibility. Because of that, it is recommended to carefully design and conduct the examinations which contain standardized characteristics such as validity, reliability, and practicality (3).

The validity of a test is composed of content validity, structure validity, and standard validity. One of the main challenges of test validity is the fulfillment of educational goals (4). Usually, MCQs that is designed according to the exam blueprint is reliable (5, 6). It is possible to design more appropriate questions and enhance their content validity by using a test blueprint. The test blueprint can be prepared in three ways: 1. An expert panel (the outlines, chapters, and percentage of questions can be agreed upon in a

group discussion with members or with peer group). 2. Task analysis & job description. 3- Inspiration from the reference books. The preparation of the test blueprint requires task analysis. The benefits of task analysis include appropriate goals, content, teaching methods, learning experiences, and evaluation (5).

Given the significance of the question analysis, the studies showed that the quality of MCQs is not consistent in terms of structure and learning difficulty. Therefore, each university has to intervene in the evaluation of MCQs. The evaluation of the quality of the MCQs in the final examination of Mazandaran University of Medical Sciences revealed that 64% of 1478 related questions to 25 exams had one or several structural defects. The majority of questions had a similar taxonomy level (7). The results also showed an unsatisfactory performance for the score of the final examination for the Isfahan University of Medical Sciences (8). The question analysis of the specialized courses at the Isfahan University of Medical Sciences was also substandard (9). Based on the study carried out by Shakoornia et al., more than half of the questions designed by the faculty members of Shapur University of Medical Sciences were correct and with a suitable structure (10). Some studies confirmed the positive influence of running empowerment workshops on the quality of MCQs (11). The importance of teacher empowerment workshops has been stated in the study of Abdulghani et al. (2017), in which the quality of MCQs was improved in terms of difficulty index, enhancing discrimination index, presence/absence of grammatical errors, and improving the Bloom's cognitive level (12). Furthermore, several other studies (12-15) have reported that the knowledge levels of the students are affected by the higher quality of MCQs. One of the most important goals of the Zanjan School of Nursing and Midwifery is to enhance the quality of the exams and the teaching-learning process. Given the undeniable value of theoretical courses in the academic term for the Zanjan School of Nursing and Midwifery students, the absence of the required information from the analysis on MCQs at the academic term and the necessity for making appropriate interventions in this regard; The present study aimed to investigate the impact of a standardization MCQs workshop on the performance and satisfaction of nursing and midwifery faculty members in designing the final examination questions and student satisfaction in Zanjan University of Medical Sciences.

Materials and Methods

This was a pre-post single-group interventional study; the sample population was the tutors of Zanjan University School of Nursing and Midwifery. The sampling population was selected by a simple random method. There are 50 faculty members at the Zanjan School of Nursing and Midwifery. Their academic rankings were six associate tutors (12%), five assistant tutors (10%), and 39 instructors (78%). The curriculum of this study was designed according to Kern's six-step model (16) of development. The six steps of this approach include problem identification and general needs assessment, targeted needs assessment, goals and objectives, educational strategies, implementation and evaluation, and feedback (Figure.1).

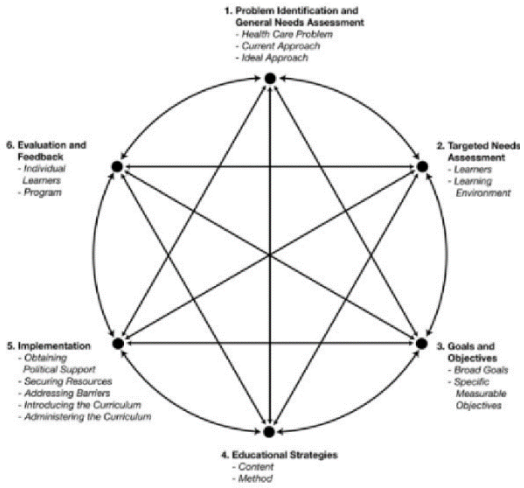

Fig 1: Steps to do the work based on kern model

In order to determine the discrimination index and needs assessment criteria, the articles were searched at the English and Persian

databases of the Scientific Information Database (SID), Magiram, Eric, Pubmed, and Google scholar. The English keywords, including Standardization, Multiple Choice Questions, and Bloom's Taxonomy and their Persian equivalent, were used to look up the articles. The review of these articles revealed that the primary educational concerns of the educational centers, such as nursing, were the quantitative and qualitative evaluation of tests and improvement of the existing standards by taking corrective measurements (1, 12-15, 17-19). The needs assessment of the Monitoring and Evaluation Committee of the Education Development Office of the School of Nursing and Midwifery indicated that the analysis of the final examination questions was vital. The examination of the MCQs at the end of the second semester of 2017-2018 determined the need for a correction. The main objects of this study include teachers empowerment for designing standard questions according to the results of structural level and taxonomy of questions by holding a "designing exam questions" workshop, evaluation of exam questions quality for all the tutors by Millman checklist and Bloom taxonomy after running a workshop, providing tutors and faculty managers with feedback on the results of the evaluation of questions, and develop sensitivity in esteemed tutors for question designing.

After presenting the evaluation report of the questions of the second semester of 2017-2018 and in the next session of the Monitoring and Evaluation Committee, the solutions were revised by brainstorming and by a panel of experts (PoE). Since the majority of the final examination was MCQs, it was decided to run an empowerment workshop on the standardization of final exam questions for the faculty members. Teacher
empowerment was performed as a single-session workshop in which the educational content consisted of familiarity with the test blueprint, principles of designing objective questions, face validity, content validity, and how to examine the difficulty/discrimination indices of questions. After each session, the PowerPoint files of the respected workshop were given to all tutors. Then, all the final examination questions were received from faculty by the end of the first semester of 2018-2019, the exam papers for each discipline were placed in a row, and one-third of the questions were evaluated randomly.

The results of the evaluation were reviewed at the third session of the Monitoring and Evaluation Committee since the previous evaluation of the second semester of 2017-2018. The first semester of 2018-2019 has addressed one-third of the MCQs; it was probable that questions with standard taxonomy structure have been excluded. Therefore, all the tutors were asked to deliver their exam papers to the esteemed expert of the Education Development Office or the executor for the analysis and assessment. Moreover, the Millman checklist and Bloom's taxonomy levels were automated to all the tutors. Both of these suggestions were applied for the evaluation of the final questions of the second semester of 2018-2019.

Finally, this approach was assessed according to Kirkpatrick's levels of reaction, learning, and behavior. The satisfaction level was evaluated for both tutors and students. A researcher-made questionnaire consisted of eight items that were used to evaluate the level of satisfaction of tutors. These items consisted of satisfaction with the content of the workshop (clarity, objectives and its content relevance to the field of activity and profession of participants), workshop design (an appropriate difficulty level of the materials and their presentation pace during the workshop), the instructor/instructors of the workshop (having the required proficiency over the materials), and the results of the workshop (fulfillment of the predefined objectives and applying what has been learned). The questions were scored on a five-point Likert scale, ranging from 1 to 5, where a high score indicated a greater level of satisfaction. Five educational experts approved the content validity of the questionnaire and its reliability was confirmed for 16 tutors with Cronbach's alpha coefficient of

0.855. The level of satisfaction was remeasured after one session of the workshop.

A researcher-made questionnaire consisted of seven items that were used to evaluate the level of satisfaction of students. These items were practicing the predefined educational goals of the lesson plan in exam questions, even distribution of questions among the lesson plan objectives, the compatibility level of quality of the questions with the actual taught material, even distribution of questions weight, the compatibility of the test type with the teaching approaches, and the appropriate exam time. The questions were scored on a five-point Likert scale. The content validity of the questionnaire was approved by eight nursing and midwifery experts (CVI=0.96 - CVR=0.93), and three students confirmed its face validity. The reliability of the questionnaire was confirmed for 30 students and with Cronbach's alpha coefficient of 0.854.

The learning of the faculty members was assessed by randomly analyzing one-third of MCQs of the final exam in one pre-intervention semester (932 questions) and for two consecutive post-intervention semesters (1679 questions) were randomly selected and analyzed. The current study analyzed and the MCQs of undergraduate and post-graduate for all the educational groups of medical-surgical nursing, community health, psychiatric, pediatrics, critical care, operating room-anesthesiology technology, and midwifery. The questions were analyzed anonymously to comply with ethical codes, and the results were reported to the tutors and the Education Development Committee of the Faculty and Zanjan University of Medical Sciences. Due to the abnormal distribution of data in the Kolmogorov–Smirnov test, data analysis was performed using descriptive statistics, Kruskal-Wallis tests, the Mann–Whitney U test, and chi-square at 95% confidence interval.

Results

A total number of 932 questions were examined in the pre-intervention semester, in which 80% of them belonged to the undergraduate and 20% to the post-graduate students. 65.8% of these questions were specific questions, 28.6% of these were basic questions, and 5.6% were general questions. From a total number of 1005 analyzed questions in the first post-intervention semester, 80.8% belonged to the undergraduate and 19.2% related to post-graduate students. 59.4% of the questions were specialized, 33.2% were basic, and 7.4% were general questions. From a total number of 674 analyzed questions in the second post-intervention semester, 98.8% belonged to undergraduate, and 1.2% were related to post-graduate students. 98.8% of these questions were specialized, and 1.2% were basic questions. According to Table.1, the frequency of the Millman checklist in all three stages of analysis (one pre-intervention semester, two post-intervention semesters) was usually over 90% most of the time.

Table 1: Frequency distribution of structural problems and structural level of multi-choice questions in the second semester of 2017-2018, first and second semester 2018-2019

semester	Before the intervention		Time 1 after the intervention		Time 2 after the intervention	
Phrase	Yes	NO	Yes	NO	Yes	NO
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
1. Is the majority of the information included in the question stem?	907 (97.3)	25 (2.7)	997 (99.2)	8 (0.8)	668 (99.1)	6 (0.9)
2. Does the question assess a specific purpose of learning?	932 (100.0)	0 (0.0)	1005 (100.0)	0 (0.0)	674 (100.0)	0 (0.0)
3- Are the words used in the stem or options clear and directly expressed?	932 (100.0)	0 (0.0)	1005 (100.0)	0 (0.0)	674 (100.0)	0 (0.0)

4- Has the use of negative option for negative stems been avoided?	926 (99.4)	6 (0.6)	1001 (99.6)	4 (0.4)	674 (100.0)	0 (0.0)
5. Has it been prevented from using options such as "all items", "none" and "combined options"?	869 (93.2)	63 (6.8)	948 (94.3)	57 (5.7)	629 (93.3)	45 (6.7)
6. Has he been prevented from using each other's conflicting options?	928 (99.6)	4 (0.4)	1005 (100.0)	0 (0.0)	674 (100.0)	0 (0.0)
7- Have positive words been used in the question area or if the question stem is negative, negative words have been specified?	904 (97.0)	28 (3.0)	973 (96.8)	32 (3.2)	659 (97.8)	15 (2.2)
8. Is this question independent of other questions?	932 (100.0)	0 (0.0)	1005 (100.0)	0 (0.0)	674 (100.0)	0 (0.0)
9- Are the options also stone in terms of length, lexical structure and writing style?	620 (66.5)	312 (33.5)	838 (83.4)	167 (16.6)	610 (90.5)	64 (9.5)
10. Have duplicate phrases been avoided in the options as much as possible?	924 (99.1)	8 (0.9)	1005 (100.0)	0 (0.0)	673 (99.9)	1 (0.1)
11- Are the words used in the stem or options correct in terms of spelling?	916 (98.3)	16 (1.7)	1004 (99.9)	1 (0.1)	671 (99.6)	3 (0.4)
12. Are options listed vertically?	282 (30.3)	650 (69.7)	362 (36.0)	643 (64.0)	248 (36.8)	426 (63.2)
13. Is it clearly a correct option?	932 (100.0)	0 (0.0)	1005 (100.0)	0 (0.0)	674 (100.0)	0 (0.0)
14. Are the stems of questions and options positive?	805 (86.4)	127 (13.6)	898 (89.4)	107 (10.6)	602 (89.3)	72 (10.7)

The lowest frequency belonged to question number 12. The percentage of scores in the first pre-intervention semester were 30.3% to 100%, in the first post-intervention semester, 36-100 percent, and in the second post-intervention semester was 36.8-100 percent.

Based on the results of the Table.2, a total of 932 MCQs was analyzed in the pre-intervention stage, which 65.8% of them had a favorable structure.

Table 2: Status Of Structural Rules And Taxonomy Level of Final Multi-Choice Questions of Zanjan Nursing And Midwifery Faculty Per Semester

Time		time					
Structure and Taxonomy		Before the intervention		Time 1 after the intervention		Time 2 after the intervention	
		N	%	N	%	N	%
Structure level of questions	Optimal level	613	65.8	826	82.2	563	83.5
	Undesirable level	319	34.2	179	17.8	111	16.5
Score the structure of questions		Mean±SD		Mean±SD		Mean±SD	
		12.67±0.86		12.99±0.75		13.06±0.69	
Kruskal-Wallis test result				X^2=110.291 P=0.000			
Taxonomy Questions	Knowledge & Comprehension	754	80.9	849	84.5	594	88.1
	Application & Analysis	167	17.9	142	14.1	73	10.8
	Synthesis. and Evaluation	11	1.2	14	1.4	7	1.0
Chi-square test result				X^2=16.59 P=0.002			

The mean and standard deviation for the score of question structure in the pre-intervention semester was 12.67±0.86. The diversity of different types of questions based on Bloom's taxonomy includes comprehension and knowledge (80.9%), analysis and application (17.9%), and synthesis and evaluation (1.2%). In the two post-intervention semesters, 1679 questions were analyzed in two consecutive periods. In this regard, the mean and standard deviation of the structural rules of questions in the first and second post-intervention semesters were 12.99±0.75 and 13.06±0.69, respectively. The result of the Kruskal-Wallis test indicated a statistically significant difference (p-value less than 0.001) and an improved mean of structural rules observed after the intervention.

Based on Bloom's taxonomy, the highest and lowest types of questions in all three semesters (one pre-intervention, two post-intervention) belonged to the pairs of comprehension-knowledge and synthesis-evaluation, respectively. The comparison of the pre-post frequency distribution taxonomy was statistically significant (p-value=0.002). According to Table.3, the satisfaction level of tutors was more than 50% in all parts of the empowerment workshops such as content, design, and instructors. Subsequently, the findings of Table.4 indicated an insignificant relationship between the questions of two post-intervention semesters in terms of even distribution of questions among the lesson plan objectives, the compatibility level of quality of the questions with the actual taught material, even distribution of questions weight, the compatibility of the test type with the teaching approaches, and the appropriate exam time. More than 50% of the students were satisfied with the quality of final exam questions in both semesters.

Table 3: Satisfaction status of tutors participating in the workshop

	Phrases	Mean±SD	Level		N	%
Workshop Content	The objectives of the workshop were clearly stated.	4.06±0.68	Workshop Content level	Optimal	9	56.2
	The content of the workshop was related to my field of activity and occupation.	4.59±0.47		Undesirable	7	43.8
Workshop design	The difficulty level of the materials presented during the workshop was appropriate.	3.94±1.06	Workshop design level	Optimal	11	68.8
	The speed of presentation during the workshop was appropriate.	4.25±0.68		Undesirable	5	31.2
Instructor / instructors workshop	The instructor / teachers of the workshop had the necessary mastery and skills to present the material.	4.06±0.57	Instructor / instructors workshop level	Optimal	9	56.2
	The teacher's behavior during the workshop was intimate and friendly.	4.50±0.63		Undesirable	7	43.8
Workshop Results	By attending the workshop, I achieved the goals specified at the beginning of the program.	4.13±0.50	Workshop Results level	Optimal	14	87.5
	I can apply what I learned in the workshop.	4.19±0.65		Undesirable	2	12.5
	Total score	33.81±3.83	Total level	Optimal	8	50
			Undesirable	Undesirable	8	50

Table 4: a- Students' satisfaction with end-of-semester questions by time

Row	Item	Time 1 after the intervention	Time 2 after the intervention	Mann-Whitney test result
		Mean ± SD	Mean ± SD	
1	Exam questions covered all the goals mentioned in the lesson plan.	3.55±1.18	3.51±1.07	Z=0.706 P=0.480
2	Distribution of questions ratio among the objectives of the lesson plan was appropriate.	3.36±1.16	3.44±1.03	Z=0.534 P=0.593
3	The content of the questions was taught in accordance with the actual contents.	3.30±1.16	3.28±1.04	Z=0.307 P=0.758
4	The exam questions were clearly adapted to the lesson plan's learned and objectives.	3.37±1.23	3.30±1.14	Z=0.571 P=0.568
5	The weighted distribution of the questions was fair according to the objectives of the lesson plan.	3.47±1.14	3.37±1.08	Z=1.025 P=0.305
6	The type of test questions was in harmony with the educational method.	3.35±1.17	3.35±1.09	Z=0.116 P=0.908
7	The duration of the exam was appropriate according to the number of questions.	3.43±1.40	3.49±1.24	Z=0.038 P=0.907
	Total	23.83±6.55	23.75±5.88	Z=0.131 P=0.896

Table 4: b- Students' satisfaction with end-of-semester questions by time

		Category Total Student Reviews			
		Time 1 after the intervention		Time 2 after the intervention	
Semester		N	%	N	%
Level of satisfaction					
Level of satisfaction	Optimal	76	55.1	85	56.3
	Undesirable	62	44.9	66	43.7

Discussion

The results of this study revealed that running the empowerment workshops on standardization of final examination MCQs had increased the performance and satisfaction levels of the tutors of the School of Nursing and Midwifery of Zanjan University of Medical Sciences. After the workshops, the structural rules of questions were increased from 65.8% to 83.5%. Similarly, the taxonomy level of questions in the areas of knowledge and comprehension has increased from 80.9% to 88.1%.

Most of the educational studies and researches focus on the structural rules of questions and designing questions with a high level of the taxonomy. If a test has a reduced taxonomy level and inappropriate structural rules, then it suffers from an unfavorable content and construct validity, negatively affects the learners' motivations, and waste the efforts of tutors and the education system (1). While the MCQs are extensively used for the evaluation of nursing students, the structural quality of questions is inconsistent (20).

Sanagoo *et al.* conducted a study (2009) that investigated the specific course questions in the educational groups of child and family nursing, community health, and psychiatric or mental health nursing. The results indicated that 5.7% and 94.3% of the questions had moderate and good structures, respectively. The taxonomy level of the questions consisted of 96.4% level one (knowledge and comprehension) and 3.6% level two (application and analysis). There was no question in level three of taxonomy (1). According to the study of Baghaei et al. in the Urmia School of Nursing and Midwifery (2015),

there were three levels of weak, moderate, and good structure with a frequency of 1.3%, 14.6%, and 84.1%, respectively. The taxonomy level of questions consisted of 84.6% level one (knowledge-comprehension), 13.5% level two (application-analysis), and 1.9% level three (synthesis-evaluation) (21). The results of the current study are in line with previous papers in terms of the structural rules of questions and taxonomy level one. Even though there is no established standard for the percentage distribution of taxonomy, but in most cases, the taxonomy level of knowledge should not exceed 50% (1, 19). Since MCQs are mainly designed for cognitive purposes and the majority of the final examination questions about undergraduate are composed of them, it is recommended to address higher cognitive levels in lesson planning and to use modern teaching methods for the development of problem-solving skills and critical thinking (22). Because designing exam questions in higher levels of taxonomy are difficult and time-consuming, it is suggested that tutors address this issue from the very first semester.

In this study, the final exam questions had a moderate structural level, which was parallel with the research of Hijji et al. (2017). The study of Hijji et al. had analyzed 98 MCQs and reported 377 grammatical mistakes; in other words, the majority of MCQs (90 items, 91.8%) had one or more mistakes and only eight questions (8.2%) followed grammatical rules of MCQs. These mistakes include verbal errors, various stem problems, and response options. It should be noted that most of the faculties did not use question analysis for assessment of the exam accuracy (23). Designing high-quality MCQs is challenging and time-consuming (23); hence by a hasty and distracted MCQs design, instructors compromise the quality of the questions (20, 24). Derakhshan et al. (2015) aimed to reflect the influence of empowerment programs for the improvement of the structural rules of questions. They conducted a three-stage empowerment program, which in the first stage, the textbooks of "Principles of Design and Evaluation of MCQs" were provided to all educational groups. All the questions of the residency promotion exam of Qazvin University of Medical Sciences in 2012 were examined per structural rules and according to the Millman checklist with 19 items. One hundred fifty questions were reviewed from each educational group. Subsequently, all of the educational groups were given the results of each field. In the third stage, educational specific workshops were held to review each group's question before the residency promotion exam of 2013. After examination of these questions, it was revealed that the highest number of structural errors of the 2012 and 2013 promotion exams was related to the negative questions, which included 178 negative items for 2012 and 176 items for 2013 (13).

The main structural errors of this study were the horizontal and non-vertical writing of the options, which contradicted the findings of Derkhshan et al. (13). The study of Meyari *et al.* (2013) has assessed the quality of questions for the residency promotion exam through the use of intervention feedback techniques in line with their technical characteristics in medical education ,such as objectivity, punctuality, directness, respectfulness, unbiased performance, and combined with improvement recommendations. It showed an improved level of taxonomy of the questions for the experienced MCQs designers. They have participated in the previous workshops and complied with the Millman checklist (14). The study conducted by Taheri at al. (2013) showed an increased number of correct questions in line with the Millman checklist and after running 12 one-day training workshops for test designers (for each educational group) (18). Their research confirmed the findings of this study.

Limitations

Due to the lack of required software, the difficulty and discrimination indices for MCQs were unspecified. Besides, this study only analyzed the MCQs, so other forms of items such as descriptive and true/false were neglected. The study team was unable to contain and address these two constraints.

Conclusion

The findings of this study indicated that the performance and satisfaction level of teachers have improved after workshops; therefore, it is suggested to hold question designing workshops for teacher empowerment continuously.

Acknowledgments

This project has been performed with the financial support of the National Center for Strategic Research in Medical Education of the Ministry of Health and Medical Education of Iran, with the project number of 972270. We like to extend our gratitude to all the tutors, instructors, experts, managers, and deputy director of education and students of the Zanjan School of Nursing and Midwifery. They kindly cooperated with us in the accomplishment of this study.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Sanagoo A, Jouybari L, Ghanbari Gorji M. Quantitative and Qualitative Analysis of Academic Achievement Tests in Golestan University of Medical Sciences. Res Med Educ. *2010;2(2):24-32.*

2. Cohen-Schotanus J. Student assessment and examination rules. Med Teach. *1999;21(3):318-21.*

3. Shabani H. Educational and training skills. Tehran: SAMT publication co. *2001*

4. Saberian M, Salemi S. (2002) How to write a lesson plan? 2nd ed, 2002 salemi Pub Co, Tehran*: 119-120. [Persian].*

5. Department of Medical Education and Development, Faculty of Nursing and Midwifery, Tehran University of Medical Sciences. Test design steps. Nursing.medilam.ac.ir website *2008. [Access date: 6/19/2018];* Available from: URL: <http://nursing.medilam.ac.ir/Portals/4/matalebdarsi/azmoon.pdf.

6. AlMahmoud T, Elzubeir MA, Shaban S, Branicki F. An Enhancement-focused framework for developing high quality single best answer multiple choice questions. Educ Health (Abingdon). *2015;28(3):194-200.*

7. Haghshenas M, Vahidshahi K, Mahmudi M, Shahbaznejad L, Parvinnejad N, Emadi A. Evaluation of multiple choice questions in the school of medicine, Mazandaran University of Medical Sciences, the first semester of 2007. Strides Dev Med Educ. *2009;5(2):120-7.*

8. Hoseini H, Rezaei Dehaghani A. Validity, Reliability and Difficulty of Multiple-Choice Questions in Specialized Courses of Final Exams in Bachelor Nursing at Isfahan University of Medical Sciences. Educ Dev Judishapur. *2019;9(2):129-36.*

9. Kazemi A, Ehsanpour S. Item Analysis of Core Theoretical Courses Exams for Midwifery Students in Isfahan University of Medical Sciences. J Iranian Journal of Medical Education. *2011;10(5).*

10. Shakoornia A, Khosravi A, Shariati A, Zarei A, editors. Survey on multiple choice questions of faculty members of Jondishapor Medical University of Ahwaz. The 8th National Congress of Medical Education Kerman: Kerman University of Medical Sciences; *2007: 44.*

11. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurs Educ. *2012;37(3):98-104.*

12. Abdulghani HM, Irshad M, Haque S, Ahmad T, Sattar K, Khalil MS. Effectiveness of longitudinal faculty development programs on MCQs items writing skills: A follow-up study. PloS one. *2017;12(10):e0185895.*

13. Derakhshan F, Allami A, Ahmadi S. Effect of Faculty Training Programs on Improving Quality of Residency Exams in 2013-2014. Res Med Educ. *2015;7(1):19-26.*

14. Meyari A, Beiglarkhani M. Improvement of design of multiple choice questions in annual residency exams by giving feedback. Strides Dev Med Educ. *2013;10(1):109-18.*

15. Ahmad RG, Hamed OA. Impact of adopting a newly developed blueprinting method and relating it to item analysis on students' performance. Med Teach. *2014;36(suppl1):S55-61.*

16. Kern DE, Bass EB, Thomas PA, Howard DM. Curriculum Development for Medical Education: A Six Step Approach: Johns Hopkins University Press; *1998.*

17. Masters JC, Hulsmeyer BS, Pike ME, Leichty K, Miller MT, Verst AL. Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. J Nurs Educ. *2001;40(1):25-32.*

18. Taheri M, khoshrang H, Asadi louyeh A, Hidarzadeh A. Quality of Residents' Promotion Exams before & after Educational Intervention in 2010-2011. Iran J Med Educ. *2013;13(7):551-60.*

19. Sayyahmelli M, Barzegar M, Bilan N, Aslanabadi S, Khoshbaten M, Ghasem zadeh A, et al. Comparison multiple-choice questions quality parameters of pediatric, general surgery, internal medicine and genecology and obstetrics residency tests between preboard examination of Tabriz University of medical sciences and national board examination in 2010 and 2011. J Med Educ Dev. *2015;8(18):43-53.*

20. O'Rae A, Hnatyshyn T, Beck AJ, Mannion C, Patel S. An Appreciative Inquiry Into Nurse Educators' Exam Practices. Nurs Educ Perspect. *2019;40(4):234-5.*

21. Baghaei R, Feizi A, Shams S, Naderi J, Rasouli D, Faculty M. Evaluation of The Nursing Students Final Exam Multiple-Choice Questions in Urmia University of Medical Sciences. Nurs Midwifery J. *2016;14(4291):291-9.*

22. Salsabili N. Using the problem solving approach in designing and developing social studies curriculum in middle school. J Curr Stud. *2006;1(3):67-104.*

23. Hijji BM. Flaws of Multiple Choice Questions in Teacher-Constructed Nursing Examinations: A Pilot Descriptive Study. J Nurs Educ. *2017;56(8):490-6.*

24. Redmond SP, Hartigan-Rogers JA, Cobbett S. High time for a change: psychometric analysis of multiple-choice questions in nursing. Int J Nurs Educ Scholarsh. 2012;9(1):ISSN (Online) *1548-923X.*

Maleki A, Ghahremani Z, Ghorbani F. Investigating the Impact of a Multiple Choice Questions Standardization Workshop on the Performance and Satisfaction of Nursing and Midwifery Faculty Members for Designing the Final Examination Questions and Student Satisfaction in Zanjan University of Medical Sciences. J Med Educ Dev. 2020; 13 (38) :59-68