ORIGINAL ARTICLE

دراسة سير كيفيه الاسئله ذو عدة اجوبه في الإمتحانات الكتبيه في جامعه بابل في الفصل الاول من عام ١٣٨٥ حتى ١٣٩٠ على اساس الهيكليه والمحتوى (نوع تاكسونومى)

التمهيد و الهدف: بعد اليوم دراسة كيفيه الاسئله ذوعده اجوبه على اساس الهيكليه و المحتوى من من الامور الإستراتيجيكيه فى الانظمه التعليميه و تعتبر من اكثر العوامل المؤثره فى ضمان كيفيه التعليم. تسعى هذه الدراسه الى تقييم معايير الإختبارات ذوعده اجوبه فى جامعه بابل فى الفصل الاول من عام ١٣٨٥ حتى ١٣٩٠ هـ.ش على اساس الهيكليه و توزيع سطح العلم.

الأسلوب : تم استعمال استماره ميلمن لاجل تقييم هيكليه الاسئله . ايضا تم اتخاذ ثلاثه مستويات للأسئله . الاول التذكر و االثانى الفهم و الثالث مورد الاستعمال .تم تحليل المعطيات عبر برنامج spss النسخه ١٦ و تم استخدام المعادلات الاحصائيه و تم تعيين ($P<0.05$)

النتائج: كان هناك ٧٠٪ من الاسئله ذو نقص من جهه الهيكليه و ذلك قبل اجراء الفحص و قد نزل هذا المستوى الى ٣٠٪ بعد اجراء التقييم و هذا الامر كان واضحا فِ المعايير الإحصائيه ($p=005.0$). كان هناك ٩.٨٣٪ من الاسئله قبل الاختبار و ٥.٧٩٪ بعد الإختبار فى مستوى تاكسونومى.

الأستنتاج: اشارت النتائج الى كيفيه هيكليه الاسئله كانت جيده لكن من حيث توزيع المستوى العلمى لم يكن هناك تغيير يذكر من مستوى تا كسونومى واحد الى جهه تاكسونومى اثنان و ثلاثه لذا نرى من الضرورى لفت نظرالمعنيين فى هذا المجال الى اقامه دورات تنمى كيفيه التعليم و التأكيد على ارتقاء مستوى الاسئله ذو عده اجوبه قسب توزيع المستوى العلمى.

الكلمات الرئيسيه: اسئله ذو عدةة اجوبه ، هيكليه ، توزيع المستوى العلمى (تاكسونومى)، جامعه بابل.

بابل یونیورسٹی کی فیکلٹی آف پیرا میڈیکل سائنس میں آبجیکٹیو سوالات کا جائزہ ۔ یہ تحقیق دوہزار سات سے دوہزار بارہ تک کے عرصے پر محیط ہے

بیک گراونڈ: آج کے نظام تعلیم میں آبجیکٹیو سوالات کا اسٹرکچر اور مضمون کے لحاظ سے جائزہ لینا ایک اہم ضرورت شمار ہوتا ہےاور اسے تعلیمی عمل کو بہتر بنانے کا ایک موثر ذریعہ سمجھا جاتا ہے۔اس تحقیق میں بابل فیکلٹی آف پیرامیڈیکل سائنس میں دوہزار سات سے دوہزار بارہ تک کے عرصے میں آبجیکٹیو سوالات کا مختلف لحاظ سے جائزہ لیا گیا۔

روش: میلمین چیک لسٹ سے سوالات کے اسٹرکچر کا جائزہ لیا گیا، سوالوں کا فہم و ادراک ، حافظے اور عمل درآمد کے لحاظ سے بھی جائزہ لیا گیا۔ ڈیٹا کا ایس پی ایس ایس سولہ سافٹ ویر سے تجزیہ کیا گیا۔

نتیجے: اسٹرکچر کے لحاظ سے ستر فیصد سوالات تحقیق سے پہلے ناقص تھے جبکہ تحقیق کے بعد ان کا نقص تیس فیصد تک رہ گیا تھا۔

سفارشات: اس تحقیق سے پتہ چلتا ہےکہ جائزے کے بعد سےسوالات میں بہتری آئي ہے لیکن ٹیکسونومی لیول کے لحاظ سے کوئي خاص فرق دکھائي نہیں دیا لھذا اس بات کی ضرورت ہے کہ تعلیمی مراکز میں ورک شاپ لگا کر سوالات کو بہتر بنانے کی کوشش کی جائے ۔

کلیدی الفاظ: آبجیکٹیو سوالات ، ٹیکسونومی ، پیرامیڈیکل کالج ۔

Evaluation of Multiple Choice Questions Quality Trend as Structure and Taxonomy

Background: Evaluation of multiple-choice questions is a strategic activity and the most effective tool in educational system and improvement. In this study, the quality of some indexes of multiple-choice exams in Babol Para Medical faculty was investigated on the basis of structure and knowledge level distribution in the first semesters of 2007 and 2012.

Methods: The Milman checklist was used for evaluating the structure quality of these questions. And these questions were also evaluated in terms of knowledge level in 3 taxonomies: 1. Reminding, 2. Perception and 3. Applying. Data analysis was performed by SPSS 16.

Results: More than 70% of these multiple-choice pre-tests contained structural flaws, while the errors decreased from 70% to 30% in post-test and this difference was statistically significant ($p= 0.05$). 83.9% and 79.5 % of these questions were located at taxonomy 1 in pre and post-tests, respectively.

Conclusion: The result showed that the structural quality of these questions had desirable improvement, but there was no significant difference in taxonomy shifts- from 1 to 2, and 3- in terms of knowledge level distribution. Therefore, it is necessary to hold workshops in order to improve the quality of multiple-choice questions in terms of knowledge level distribution at University.

Keywords: Multiple-choice Questions, Structure, Knowledge Level Distribution, Para Medical Faculty

بررسی روند کیفیت پرسش های چند گزینه ای آزمون های کتبی دانشکده پیراپزشکی بابل در نیمسال اول ۱۳۸۵ تا ۱۳۹۰ بر حسب ساختار و محتوا(نوع تاکسونومی)

زمینه و هدف: امروزه فرآیند بررسی پرسش های چند گزینه ای بر حسب ساختار و محتوا، به عنوان یکی از فعالیت های استراتژیک در سیستم های آموزشی مطرح بوده و از مؤثرترین ابزارهای تضمین ارائه کیفیت بهینه آموزش در نظر گرفته می شود. در مطالعه حاضر، کیفیت برخی از شاخص های آزمون های چند گزینه ای دانشکده پیراپزشکی دانشگاه علوم پزشکی بابل در دو نیمسال اول ۱۳۸۵ و ۱۳۹۰ برحسب ساختار و توزیع سطح دانش مورد بررسی قرار گرفت.

روش: از چک لیست میلمن جهت بررسی کیفیت ساختاری سئوالات استفاده شد. همچنین پرسش های مزبور از لحاظ سطح دانش در سه تراز یادآوری، فهم و درک و کاربرد مورد بررسی قرار گرفتند. تجزیه و تحلیل آماری داده ها به وسیله نرم افزار SPSS نسخه ۱۶ و با استفاده از شاخص های مرکزی و پراکندگی و آزمون مجذور کا انجام گردید. سطح معنی داری در این پژوهش با $P<0.05$ در نظر گرفته شد.

یافته ها: از نظر ساختار، بیش از ۷۰ درصد از پرسش ها در قبل از پیش آزمون دچار نقص بوده اند، در حالی که این مقدار در بعد از اجرای برنامه جامع ارزیابی پرسش ها به کمتر از ۳۰ درصد کاهش یافته که این تفاوت از لحاظ آماری نیز کاملا معنی دار می باشد($P= ۰/۰۰۵$). ۸۳/۹ درصد و ۷۹/۵ درصد از پرسش ها به ترتیب در مراحل پیش آزمون و پس آزمون در سطح تاکسونومی یک قرار داشته اند.

نتیجه گیری: نتایج بدست آمده نشان می دهد که کیفیت ساختاری پرسش ها بهبود مطلوبی داشته اما از لحاظ توزیع سطح دانش از تراز تاکسونومی یک به سمت تاکسونومی های دو و سه تغییر قابل ملاحظه ای مشاهده نمی گردد. بنابر این ضرورت دارد که مدیران مراکز توسعه آموزش در دانشگاه به برگزاری کارگاه های ارتقای کیفیت آموزش با تاکید ویژه بر ارتقای کیفیت پرسش های چند گزینه ای بر حسب توزیع سطح دانش مبادرت نمایند.

واژه‌های کلیدی: پرسش‌های چندگزینه‌ای، ساختار، توزیع سطح دانش (تاکسونومی)، دانشکده پیراپزشکی بابل

Abdoliman Amouei[1], Reyhaneh Barari[2], Dariush Naghipour[3], Yousef Mortazavi[4*], Seyed Reza Hosseini[5]

[1] Environmental Health Engineering Department, Para Medicine and Health Faculty, Babol University of Medical Sciences, Babol, IRAN
[2] Allameh Tabatabaei University, Tehran, IRAN
[3] Environmental Health Engineering Department, Health Faculty, Guilan University of Medical Sciences, Rasht, IRAN
[4] Anesthesias & Operating Room Department, Para Medicine and Health Faculty, Babol University of Medical Sciences, Babol, IRAN
[5] Social Medicine Department, Medicine Faculty, Babol University of Medical Sciences, Babol, IRAN

[*] Anaesthesia & Operating Room Department, Para Medicine and Health Faculty, Babol University of Medical Sciences, Babol, IRAN

Tel: +98-1112234366
Fax: +98-1112234367
E-mail: imnamou@yahoo.com
Received: May 26, 2013
Accepted: Apr 20, 2014

INTRODUCTION

Nowadays, it is important to investigate and evaluate questions of a test that is a strategic activity in educational system. It is also considered as the most effective guarantee tool in teaching (1, 2, and 3). Therefore, in order to develop educational systems quality, it is absolutely necessary to examine and analyze the exams accurately. Although multiple choice question is the most common exam in medical educational system, because of easiness in performance (in large scale), high objectivity and reliability, there are many concerns about the quality of MCQ exam on its validity, content, formation, knowledge level and structure principles (4,5,6,7,8). If an exam does not have a suitable design in terms of selecting questions' taxonomy and regarding structural rules, the content validity and structure will not be satisfying. In this case, not only the main role of exam as a supplementary and important component in educational cycle is disappeared, but also it has negative effects on learners' motivations and leads to waste of human resources and educational system (6, 9). As mentioned above, considering the question's taxonomy in exam design is very basic and important. It is really as a criterion for knowledge depth and level of examines question taxonomy divided into 3 taxonomies 1- reminding simple subjects. 2- Analysis and deep perception of subjects 3- solve the problems and apply the information in this research, taxonomy 1 belonged to low taxonomy questions but taxonomies 2 and 3 were questions with high taxonomy.

Many researches have been conducted on multiple choice questions in different majors of medical sciences. Mackorbi et al at Bristol University and Harvard et al from Hampton University said that a high percentage of the multiple choice questions had structural problems and also more than 90 percent of them belonged to low taxonomy (10, 11, 12). According to Kaveh and his colleagues, the main part of evaluating criteria in multiple choice questions was falsified (13). However, In Shakurnia et al research, the questions lacked structural problems and they were according to Milman checklist indexes (14). There were many researches on evaluating the quality of multiple choice questions' structures on the basis of Milman checklist but few researches have been performed on assessing the quality of these questions in terms of taxonomy and knowledge level. As Haqshenas et al mentioned 77.7 percent of these questions were located in taxonomy 1 and the rest belonged to taxonomy 2 and 3 in terms of their quality. 46 percent of their structures were correct and the rest has one or more flaws in their structures (15). Vahidshahi et al studied the quality of some multiple choice question indexes in written exam of medical specialized scholarship in 2007, 2008 and 2009 (16).

On the basis of Milman checklist, 57.5%, 64% and 64% of these questions had few flaws in structures in 2007, 2008 and 2009, respectively. In mentioned research, 38.7%, 45% and 56% of these questions were located in 2 and 3 taxonomy in 2007, 2008 and2009, respectively. (16). The present research studied the quality of some multiple choice question indexes in Para medical college of Babol university of medical sciences during the two first semesters from 2006 to 2011, regarding the structural rules and knowledge level distribution.

METHODS

This interventional study was carried out to evaluate structural quality and 526 taxonomy of multiple choice questions derived from 12 final written exams, which were guided by 12 faculty members of Para Medical College of Medical Sciences University of Babol. The criteria included full- time job professors and their teaching in the basic and technical lessons at the university. This study had three stages including pre-test in the first semester of 2007, interventional programs like educational workshops about the evaluation of questions during 2008 to 2011 and post-test to determine the effect of educational programs on the questions quality in the first semester of 2012. There are different ways to evaluate multiple choice questions (15). The most common way is using Milman checklist with especial indexes for evaluating the quality of multiple choice questions in terms of its structure (14, 15, and 16). 12 items of Milman checklist were selected to evaluate the structural quality of multiple-choice questions in this research (Table 1).

Table 1. Milman checklist

Subjects	Yes	No
1. Is the most information located in the trunk of question?		
2. Has a question evaluated the special education purpose?		
3. Are the vocabularies in trunk and choices of questions direct and clear?		
4. Is it avoided to use negative choice for negative question?		
5. Is it avoided to use the choices such as: None of them, All of them and combination of choices?		
6. Is it avoided to use conflicting choices?		
7. Are the positive vocabularies used in the trunk of question or if they are negative, are the negative vocabularies determined?		
8. Is each question related to the other questions or not?		
9. Are the choices in the balance in terms of length, structure and style?		
10. Is it avoided to repeat the same choices?		
11. Are the words used in the trunk and choices of question correct in terms of spelling?		
12. Are the choices located vertically?		

These questions were also evaluated in terms of knowledge level in 3 taxonomies: 1- Reminding 2- Perception and 3- Applying. The validity of Milman checklist according to world standards was confirmed. The reliability of question was also proved by pretesting, evaluating, repointing the questions and determining the coefficient of correlation r= 0.89. All questions were according to correct structures, Milman checklist and the knowledge level distribution (the type of taxonomy) were evaluated by the members of medical educational development center and faculty members of mentioned courses. In this research, the name of all lessons was coded and the results were given to the related teachers confidentially. Analysis of the data, central and scattering indices determination were done by SPSS version 16. The significant level was considered $p< 0.05$ in this research.

RESULTS

The quality and validity of questions' structure were determined on the basis of Milman index (table 2) and questions' content were also designated on the basis of taxonomy (table 3).

Less than 30% of questions (29.2± 11.6%) had no flaws in pre-test in terms of structural quality, while more than 70% of questions (73.1± 4.5%) in post-test were corrected. The

Table 2. Evaluating the structural quality percentages of multiple- choice questions in written form at the first semester in 2006 and 2011 on the basis of Milman indexes.

Question quality / Lesson code	Free error		1 error		2 errors		3 and more than 3 errors		P value
	Before	After	Before	After	Before	After	Before	After	
L_1	30±11.5	96.3±1.5	20±5	3.7±0.5	10±2	0	40±4.8	0	0.001
L_2	30.8±8.3	80±5	38.5±6.5	20±1	15.4±2.2	0	15.4±1	0	0.002
L_3	16.7±9.4	100	61.1±12	0	22.2±2.5	0	0	0	0.001
L_4	29.2±17	83.9±8.6	29.2±5	12.9±1	41.7±3	3.2±1	0	0	0.005
L_5	22.2±13.4	75±5.5	11.1±4.5	25±2.3	44.4±3.3	0	22.2±2	0	0.003
L_6	30.8±15.5	83.3±6	38.5±6	16.7±1.2	30.8±2	0	0	0	0.005
L_7	31.3±14	38.3±1.5	18.75±3	40±3	37.5±1	26.7±1	12.5±1	0	0.03
L_8	23.1±10.5	68.7±6.2	46.2±7	33.3±2.4	15.4±2.8	0	15.4±1.5	0	0.01
L_9	32.5±13.8	80±8	35±6.3	15±1.6	30±2.4	5±0.8	2.5±0.5	0	0.01
L_{10}	34.3±7.5	53.3±2	37.1±4	43.3±1	22.9±2	3.3±0.6	5.7±0.8	0	0.02
L_{11}	43.8±5.8	60±5	28.1±2	33.3±1.8	15.6±1.8	3.3±0.5	12.5±1	3.3±0.5	0.03
L_{12}	25.8±12.5	65.6±4.5	45.2±3	31.2±2	25.8±3	3.1±0.7	3.2±1	0	0.02
Mean	29.2±11.6	73.1±4.5	34.1±5.4	22.9±1.5	26±2.2	3.7±0.4	10.8±1.1	0.3±0.06	0.005

Table 3. Evaluating the structural quality percentages of multiple-choice questions in written form at the first semester in 2006 and 2011 on the basis of taxonomy type.

Question quality / Lesson code	Taxonomy 1		P value	Taxonomies 2 and 3		P value
	Before	After		Before	After	
L_1	70±14.5	66.7±5.8	0.06	30±5	33.3±5.5	0.09
L_2	100	66.7±15	0.03	0	33.3±2	0.02
L_3	77.8±16	76.2±14.5	0.1	22.2±5	23.8±3.8	0.09
L_4	95.8±3	90.3±2	0.06	4.2±2.4	9.7±4.5	0.06
L_5	66.7±13	62.5±11.8	0.07	33.3±11.8	37.5±9.5	0.06
L_6	100	100	----	0	0	----
L_7	93.8±5	93.3±4.3	0.6	6.2±2	6.7±1.5	0.7
L_8	84.6±6.8	100	0.04	15.4±8	0	0.04
L_9	75±8.5	60±7.8	0.05	25±7	40±3.6	0.04
L_{10}	68.6±4.6	70±4.2	0.08	31.4±6.5	30±4.5	0.1
L_{11}	75±8.5	83.3±6.7	0.05	25±6.5	16.7±3.8	0.06
L_{12}	100	96.9±1.6	0.08	0	3.1±0.2	0.08
Mean	83.9±6.1	79.5±5.6	0.07	16.1±4.5	20.5±3.2	0.08

difference was statistically significant (p= 0.005). Of all evaluated questions, on average more than 30% of questions had no error, 34% of questions had one error, 26% contained 2 errors and the rest had more than 2 flaws in pre-test. As in post-test, 73.1% of these questions lacked errors, 4% contained 2 and more than 2 errors and the rest had only one structural error. The quality of questions was evaluated by knowledge level distribution in pre-test research. Results showed that more than three-quarters fraction of the post and pre-test questions were 83.9% and 79.5%, respectively, and located in taxonomy 1. And data analysis by SPSS indicated that there was no significant difference between pre-test and post-test in taxonomy 1.

DISCUSSION

More than 70% of pre-test questions contained errors in terms of quality while less than 30% of these multiple-choice questions had flaws in post-test and this difference was statistically significant as a whole (p = 0.005). The results showed that desirable effect of EDC (Educational Development Center) of university and EDO (Educational Development Office) of Paramedical faculty on the design and structure of exam questions is to progress and improve the faculty members of this college. Haqhshenas et al evaluated multiple-choice exams in medical college of Mazandaran university of medical sciences at the first semester of 2006 - 2007. They concluded that 46% of these questions were free of structural errors and the rest of them contained one or more structural flaws (15). The low percentage of free-error questions in terms of structure could be due to the inactivity of educational development center of mentioned university during their studies. Evaluating structural principles and taxonomy of medical specialized scholarship exam were performed by Vahid Shahi et al who concluded that 57.5% and 64% of these questions were without structural error on the basis of Milman checklist in 2007 and 2008, respectively (16). Meyari et al studied the effect of educational interferences on optimizing the multiple- choice questions design in progressing of assistant exam in schools of Dentistry in 2008 and 2009 (17). They found that 63.1% and 76.3% of these questions were without structural errors in 2008 and 2009, respectively (11). And their results were according to the results of the present research (Haqhshenas et al, Vahid Shahi et al, Meyari et al). And this research showed that the quality of questions' structure was better than before in medical education in university and science centers in Iran. More than three-forth of the pre and post test questions were 83.9% and 79.5% respectively, and located at the reminding level (Taxonomy 1). According to Haqhshenas et al 77.4% of the studied questions were categorized at taxonomy 1(15). And their results also tallied with the results of the present research. In Vahid Shahi et al. research, 38.4% and 45.6% of medical assistant questions belonged to taxonomy 2 and 3 in 2007 and 2008 (10). The rate of designed questions related to dental assistant exams in taxonomy 2 and 3 significantly increased in 2009 rather than 2008, in Meyari et al research (17). Although the recent findings were in accord with the results of present research, these results were not statistically significant in the present research.

Conclusion: The results showed that the improvement trend of structural quality had optimal growth. In spite of evaluating the questions by educational groups, there was no considerable improvement in the content quality of these questions to change higher taxonomies from taxonomy 1. It means that the examiners focus on multiple-choice questions which will be evaluated only parrot knowledge of students. Continuing this trend, the students tend to parrot learning and gain shallow knowledge. This is a basic and important issue in medical education rather other sciences. Therefore, it is necessary to improve the quality of multiple-choice questions on the basis of knowledge level distribution, and to change the taxonomy from 1 to higher taxonomies by educational development centers and other responsible persons. The results of the study were affected by limitations such as absence of faculty members of this college, prolongation of re-assessment process and other factors.

ACKNOWLEDGMENT

We are extremely grateful to the members of educational development center of medical sciences, scientific members and education staff of Para medicine faculty of Babol University of medical sciences.

REFERENCES

1. Epstein RM. Assessment in medical education. New England J Med 2007; 356(4): 387-96.
2. Friedman Ben DM. Standard setting in student assessment. Med Teach 2000; 22(2): 120-30.
3. Bazargan A. Educational assessment. Tehran: SAMT; 2011: 119-26. [Persian].
4. Seif A. Educational measurement, assessment, and evaluation. 9th ed. Tehran: Agah; 2011: 70 -80. [In Persian].
5. Seif A. Educational measurement, assessment and evaluation. 4th ed. Tehran: Doran; 2004: 426-8. [In Persian].
6. Downing SM. Reliability on the reproduce ability of assessment data. Med Educ 2004; 8(9): 1006-12.
7. Collins J. Education techniques for long learning: Writing multiple choice question for continuing medical education activities and self assessment modules. Radio Graphics 2006; 26(2): 543-51.
8. Simkin MG, Kuechler WL. Multiple choice test and student understanding: What is the connection? Decision Sci J Innovate Educ 2005; 3(1): 73-98.
9. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice question for use in nursing research and education. Collegian 2005; 12(1): 14-24.
10. Mollahadi M. Evaluating by multi choice tests. Educational strategies journal 2010; 2(4): 177-82. [Persian].
11. McCourbie P. Improving the fairness of multiple choice questions: A literature review. Med Teach 2004; 26(8): 709-12.
12. Hammond EJ, McIndoe AK, Sansome AJ, Spargo PM. Multiple-choice examinations: Adopting on evidence-based approach to exam technique. Anesthesia 1998; 53(11): 1105-8.

13. Kaveh M. Analytical assessment of multiple-choice test at medical school. Medical journal of Mashhad University of Medical Sciences 2002; 76: 89-95. [Persian].
14. Shakoornia A, Khosravi A, Shariati A, Zarei A. Survey on multiple choice questions of faculty members of Jondi Shapor Medical University of Ahwaz. Proceeding of the 8th National Congress of Medical Education. Kerman: Kerman University of Medical Sciences; 2007: 44. [In Persian].
15. Haghshenas M, Vahidshahi K, Mahmoudi M, Shahbaznejad L, Parvinnejad N, Emadi A. Evaluation of multiple choice question in the school of medicine, Mazandaran University of Medical sciences 2007. Strides in development of medical education 2008; 5(2): 126-33. [Persian].
16. Vahidshahi K, Mohagheghi M, Shakeri S, Sabouri M, Razavi SM, Mohammadi M, et al. Comparison of multiple-choice question
17. tests indices in residence diploma. [cited 2012]. Available from: URL; www.behdasht.gov.ir [In Persian].
18. Meyari A, Biglarkhani M, Zandi M, Vahedi Mohammad, Mir Esmaeili AF. The effect of education on improvement of multiple choice questions' designing in annual residency exams of dental school. Iranian journal of medical education 2012; 12(1): 36-45. [Persian].