مقاله تحقیقی

مجلهٔ دانشگاه علوم پزشکی بابل
دوره دوازدهم، ویژه نامه ۱ (آموزش پزشکی)، تابستان ۱۳۸۹، صفحه ۶۹-۶۵

تاثیر خودارزیابی سوالات امتحانی توسط گروههای آموزشی بر کیفیت آزمونهای پایان ترم

حوریه محمدی(MD) [۱]، یحیی جوادیان(PhD)* [۲]، نوین نیک بخش(MD) [۳]، مصطفی جوانیان(MD) [۴]، زینب جهانیان(MD) [۱]

۱- مرکز توسعه و مطالعات آموزش پزشکی علوم دانشگاه علوم پزشکی بابل
۲- گروه فیزیوتراپی دانشگاه علوم پزشکی بابل
۳- گروه جراحی دانشگاه علوم پزشکی بابل
۴- مرکز تحقیقات بیماریهای عفونی و گرمسیری دانشگاه علوم پزشکی بابل

دریافت: ۸۸/۸/۲۰ ، اصلاح: ۸۸/۹/۱۸، پذیرش: ۸۸/۱۲/۱۹

خلاصه

سابقه و هدف: شرکت فعال گروههای آموزشی در تجزیه و تحلیل آزمونهای پایان ترم تاثیر بسزایی در ارتقاء کیفی سوالات دارد. این مطالعه به منظور تاثیر این خود ارزیابی توسط گروههای آموزشی بر کیفیت سوالات در دانشکده دندانپزشکی دانشگاه علوم پزشکی بابل انجام شد.

مواد و روشها: این مطالعه مقطعی بر روی ۲۲ نفر از اعضای هیأت علمی دانشکده دندانپزشکی دانشگاه علوم پزشکی بابل که کارگاه ارزیابی سوالات را گذرانیده بودند، انجام شد. از آنها خواسته شد که سوالات امتحانی را قبل و بعد از برگزاری هر آزمون ارزیابی نمایند. سوالات امتحانی نیم سال دوم سال تحصیلی ۸۶-۸۵ بعنوان پره تست و سوالات نیم سال دوم سال تحصیلی ۸۸-۸۷ بعنوان پست تست مورد ارزیابی و مقایسه قرار گرفت.

یافته ها: در مجموع ۳۶۷ سوال به عنوان پره تست و ۲۸۴ سوال به عنوان پست تست مورد ارزیابی قرار گرفت. میانگین نمره کسب شده توسط اساتید (تعداد خطاها به تعداد کل سوالات) در پره تست، ۰/۲۳±۰/۲۹ با دامنه (۰/۸۷-۰) و در پست تست ۰/۱۴±۰/۱۸ با دامنه (۰/۵-۰) بود. در ارزیابی درجه تاکسونومی ۹۴٪ سوالات سال ۸۶ دارای تاکسونومی ۱و ۵/۷٪ دارای تاکسونومی ۲ و ۰/۳٪ دارای تاکسونومی۳ بود و در سال ۸۸، ۹۵/۴٪ دارای تاکسونومی ۱ و بقیه دارای تاکسونومی۲ بودند.

نتیجه گیری: نتایج مطالعه نشان داد که بررسی و ارزیابی سوالات امتحانی قبل از انجام امتحانات در داخل گروه های آموزشی در افزایش کیفیت سوالات امتحانی موثر می باشد.

واژه های کلیدی: ***خودارزیابی، سئوالات امتحانی، گروههای آموزشی، آزمونهای پایان ترم.***

مقدمه

یکی از مهمترین وظایف اعضای هیات علمی ارزیابی دانشجو می باشد که نیاز به مهارت های لازم دارد و به عنوان یک شاخص کلیدی در فرآیند یاددهی - یادگیری و تصمیم گیری برای ارتقاء یا رد دانشجویان به کار می رود (۱). بدیهی است تهیه آزمون با نوشتن سوالات پایان نمی یابد بلکه پس از آن باید به دنبال شواهدی برای اطمینان از کیفیت مناسب سوالات بود (۲). در تجزیه و تحلیل سوالات اصولا اندازه گیری دو جنبه کیفی و کمی آن مورد توجه قرار می گیرد، در ارزیابی کیفی سوال، شکل، روش تهیه، دقت در ساختن متن، سطح سوال در حیطه شناختی و چینش گزینه های درست و نادرست ارزیابی می شود و از لحاظ کمی مواردی مانند میزان اعتبار آزمون، قدرت تشخیص هر سوال و تاثیر آن در جدا ساختن افراد قوی از ضعیف، درجه دشواری هر سوال و میزان مشابهت گزینه های نادرست هر سوال و قدرت آن در جلب توجه آزمودنی مورد بررسی قرار می گیرد (۳). هر آزمون به عنوان ابزار سنجش و اندازه گیری باید دارای روایی و پایایی کافی برای سنجش صفت مورد نظر باشد (۴). بدیهی است اگر یک آزمون از طراحی مطلوب در زمینه انتخاب تاکسونومی سؤالات و رعایت قواعد ساختاری، برخوردار نباشد نه تنها نقش اصلی آزمون به عنوان جزء تکمیل کننده و حیاتی چرخه آموزش از دست می رود، بلکه این امر

* مسئول مقاله:
آدرس: بابل، دانشگاه علوم پزشکی بابل، گروه فیزیوتراپی، تلفن: ۶-۲۱۹۹۵۹۲-۰۱۱۱.
e-mail:javad835@yahoo.com

[DOR: 20.1001.1.15614107.1389.12.5.10.0]

بر فراگیران اثرات منفی گذاشته و خطر به هدر رفتن تلاش استادان و سیستم آموزشی را خواهد داشت (۵). متاسفانه بسیاری از آزمونها استاندارد نبوده و با اشکالات متعددی همراه هستند (۴). مطالعه Kehoe و همکاران نشان داد که تجزیه و تحلیل سوالات تاثیر بسزایی در ارتقاء کیفی سوالاتی دارد که مجددا در آزمون های بعدی استفاده می شوند، همچنین برای حذف سوالات مبهم و دو پهلو نیز مفید می باشد و در مجموع باعث افزایش مهارت مدرسان در طراحی سوالات آزمون و تعیین حیطه های اختصاصی محتوای یک دوره که نیازمند تاکید و شفافیت بیشتر است، می گردد (۶).

با توجه به تحقیقات انجام شده در دانشگاههای علوم پزشکی، ارزشیابی امتحانات به عمل آمده از دانشجویان که عمدتا با استفاده از سوالات چند گزینه ای انجام می گیرند، نقص های قابل ملاحظه ای دارند (۷). Saleml و همکاران معتقدند مهارت در طراحی سوال استاندارد در حال حاضر به عنوان یک مساله قابل بحث آموزشی مطرح است (۸). مطالعه Ashraf pour و همکاران نشان داد که بیش از نیمی از سوالات امتحانات پایان ترم رشته پرستاری از کیفیت مطلوب برخوردار نبودند (۹). بنابراین سوالات هر آزمون باید قبل از برگزاری، توسط مکانیزم هایی از نظر ساختاری و محتوای علمی و همچنین پس از امتحان از نظر ضریب دشواری و قدرت تمیز مورد ارزیابی قرار گیرند. یکی از این مکانیزمها، ارزیابی سوالات امتحانی توسط خود گروههای آموزشی می باشد، چون شناخت کافی نسبت به امتحانات مربوطه دارند، لذا این مطالعه به منظور تاثیر این خود ارزیابی توسط گروههای آموزشی بر کیفیت سوالات در دانشکده دندانپزشکی دانشگاه علوم پزشکی بابل انجام شد.

مواد و روشها

این مطالعه مقطعی بر روی سوالات امتحانی ۲۲ نفر از اعضای هیأت علمی گروه های اندودنتیکس، جراحی و پاتولوژی، پریو، پروتز، اطفال، رادیولوژی و بیماریهای دهان و دندان دانشکده دندانپزشکی دانشگاه علوم پزشکی بابل که کارگاه ارزیابی سوالات را گذرانیده بودند، انجام شد. از اعضای هیأت علمی گروههای آموزشی مورد مطالعه خواسته شد که سوالات امتحانی گروه خود را قبل و بعد از برگزاری هر آزمون ارزیابی نمایند. سوالات امتحانی نیم سال دوم سال تحصیلی ۸۶-۸۵ (قبل از اجرای این طرح) به عنوان پره تست و نیم سال دوم سال تحصیلی ۸۸-۸۷ (پس از ۲ سال از اجرای آن) را به عنوان پست تست مورد ارزیابی و مقایسه قرار گرفتند. مدرسین جدیدی که در این فاصله زمانی به عضویت گروه در آمده بودند وارد مطالعه نشدند. برای جلوگیری از خطاهای انسانی احتمالی و افزایش پایایی آزمون، هر سوال توسط ۲ نفر به صورت جداگانه و براساس معیارهای ۲۱ گانه millman ارزیابی گردید (۱۰).

ارزش کلیه معیارهای پرسشنامه یکسان در نظر گرفته شد و نسبت تعداد خطاهای هر مدرس به تعداد سوالات وی به عنوان نمره مدرس در طراحی آن آزمون در نظر گرفته شد. سپس سطح سوالات بر اساس سطوح شناختی بلوم در سه سطح تاکسونومی۱ (مبتنی بر حافظه)، تاکسونومی۲ (مبتنی برفهم) و تاکسونومی۳ (مبتنی بر کاربرد مفاهیم) مشخص گردید. اطلاعات بدست آمده با استفاده از آزمون Paired T-Test تجزیه و تحلیل و $p<۰/۰۵$ معنی دار در نظر گرفته شد.

یافته ها

در این مطالعه در مجموع ۳۶۷ سوال به عنوان پره تست و ۲۸۴ سوال به عنوان پست تست مورد ارزیابی قرار گرفت. میانگین نمره کسب شده توسط اساتید (تعداد خطاها به تعداد کل سوالات) در پره تست ۰/۲۳±۰/۲۹ با دامنه (۰/۸۷-۰) و در پست تست ۰/۱۴±۰/۱۸ با دامنه (۰/۵-۰) بود که این اختلاف معنی داری نبود. ٪۷۳ سوالات پره تست بدون اشکال، ٪۲۵/۶ دارای یک اشکال و ٪۱/۳ دارای ۲ اشکال ساختاری بودند در حالیکه ٪۸۲/۳ سوالات پست تست بدون اشکال و ٪۱۷/۶ یک اشکال داشته و هیچ سوالی دارای ۲ اشکال نبود (جدول ۱). بیشترین خطاها در پره تست به ترتیب مربوط به آیتم های "تناسب گزینه ها از نظر طول و پیچیدگی"با ٪۳۳" عدم استفاده از کلماتی مانند همه، هیچکدام و ..." با ٪۱۳/۵، "عدم اشکال از نظر دستور زبان در پایه و گزینه های سوال" با ٪۱۰/۶ بود. در آزمون پست تست موارد " قرار دادن سوالات و گزینه ها در یک صفحه" با ٪۳۰، "عدم اشکال از نظر دستور زبان در پایه و گزینه های سوال" با ٪۲۸ و "عدم استفاده از عبارات منفی و مشخص کردن آنها در صورت استفاده" با ٪۱۸ بیشترین خطاها بود. در ارزیابی درجه تاکسونومی ٪۹۴ سوالات سال ۸۶ دارای تاکسونومی ۱و ٪۵/۷ دارای تاکسونومی ۲ و ٪۰/۳ دارای تاکسونومی۳ بود. ٪۹۵/۴ سوالات سال ۸۸، دارای تاکسونومی ۱ و بقیه دارای تاکسونومی۲ بودند.

جدول ۱. درصد اشکالات ساختاری سوالات بر حسب گروههای آموزشی

تعداد اشکال	پره تست			پست تست		
گروه	ندارد	یک	دو	ندارد	یک	دو
رادیولوژی و بیماری های دهان و دندان	٪۷۷/۳	٪۲۲/۷	۰	٪۸۶/۱	٪۱۳/۹	۰
اندودنتیکس	٪۵۴	٪۴۴	٪۲	٪۸۰	٪۲۰	۰
جراحی و پاتولوژی	٪۷۹/۹	٪۱۸/۱	٪۲	٪۸۶/۵	٪۱۳/۵	۰
اطفال	٪۸۳/۸	٪۱۳/۹	٪۲/۳	٪۸۷/۵	٪۱۲/۵	۰
پریو و پروتز	٪۶۸/۶	٪۳۰/۵	٪۰/۹	٪۷۶/۱	٪۲۳/۹	۰

بحث و نتیجه گیری

نتایج این مطالعه نشان داد که خود ارزیابی سوالات امتحانی توسط گروه های آموزشی در کاهش اشکالات ساختاری آزمونها موثر است. اکثریت سوالات مورد بررسی در این مطالعه بدون اشکال و تعداد محدودی دارای ۱ یا ۲ اشکال بودند که در پست تست تعداد این اشکالات کمتر نیز شد که با مطالعه Shakoornia و همکاران، که بیش از نیمی از سوالات مورد بررسی توسط آنها

[DOR: 20.1001.1.15614107.1389.12.5.10.0]

فاقد اشکال ساختاری بودند، همخوانی دارد (۱۱). در مطالعه Shaban و همکاران میزان سوالات بدون اشکال، بالای ۹۵٪ گزارش شده است (۷) در حالیکه در مطالعه Haghshenas و همکاران، ۴۶٪ سوالات فاقد اشکال و در مطالعه Moattari و همکاران ۸۶٪ سوالات مستلزم اصلاح، تشخیص داده شدند (۱۲و۵).

نتایج مطالعه همچنین نشان داد که شایعترین اشکال ساختاری موجود در سوالات، "تناسب گزینه ها از نظر طول و پیچیدگی" بود که از این نظر با مطالعه Haghshenas و همکاران و از نظر "عدم استفاده از عبارات منفی و مشخص کردن آنها در صورت استفاده" با مطالعه Javadi و همکاران همخوانی داشت (۱۳و۵) ولی در مطالعه Shakoorinia و همکاران و Daripoor و همکاران دیگر موارد، بیشتر بود (۱۴و۱۱). از نظر توزیع سوالات در سطح دانش نیز نتایج بدست آمده با مطالعه Rasolinejad و همکاران که اکثریت سوالات آزمون ارتقاء دستیاران آن دانشکده تاکسونومی ۱ بودند و مطالعه Haghshenas و همکاران که بیش از ۹۰٪ سوالات علوم پایه دانشکده پزشکی در سطح تاکسونومی ۱ بودند، همخوانی دارد و این در حالی است که در مطالعه Haghshenas ۷۰٪ سوالات گروههای بالینی در سطح تاکسونومی ۱ قرار داشتند (۱۵و۵). از آنجایی که گروههای مورد بررسی در این مطالعه، گروههای بالینی دانشکده دندانپزشکی بودند جا دارد توجه ویژه ای به کاربرد و فهم دقیق مطالب شود. بررسی مطالعات انجام شده در این زمینه، یافته های مشابه مبنی بر غلبه کاربرد تاکسونومی ۱ در آزمونهای پیشرفت تحصیلی را نشان می دهند اما در مطالعات جدیدتر سالهای اخیر، سهم بیشتری از سوالات به سطوح عمیق تر دانش (تاکسونومی ۳و ۲) اختصاص یافته اند (۵). به هر حال این واقعیت که تمرکز عمده طراحان سوال بر محفوظات دانشجویان و نه فهم عمیق و تحلیل و کاربرد مفاهیم بوده است جای تامل دارد چرا که بدیهی است چنین رویکردی موجب سوق دادن دانشجویان به محفوظات و برداشت سطحی مطالب می گردد که در مورد مطالب پزشکی این شیوه یادگیری نه تنها از ماندگاری لازم در ذهن دانشجو برخوردار نیست بلکه از کارکرد لازم در زمینه پایه ریزی بستر دانش آینده و گسترش توانایی های ذهنی دانشجو کاملا دور است پس توجه ویژه به طرح سوال در سطوح عمیق تر دانش ضروری می باشد. نتایج مطالعه Zolfaghari و همکاران نشان داد که طراحان سوال به طور مستمر باید تجزیه و تحلیل سوالات را معیار کیفیت قرار دهند، سوالات غیر قابل قبول را حذف نمایند و سوالات بهتر، با ملاک های مطلوب را جایگزین و نگهداری کنند (۲).

نتایج این مطالعه همچنین نشان داد که ارزیابی سوالات امتحانی توسط گروههای آموزشی مربوطه، می تواند در اصلاح اشکالات و ایراداتی که در طراحی سوالات امتحانی وجود دارد، موثر باشد اما در حد انتظار خود را نشان نداد که شاید یکی از دلایل آن عدم مشارکت موثر گروههای آموزشی در ارزیابی اولیه سوالات از نظر کیفی و ساختاری قبل از برگزاری آزمون و همچنین ارزیابی مجدد پست تست باشد. لذا شرکت فعال تر اساتید گروههای آموزشی در بررسی و ارزیابی سوالات امتحانی اعضای گروه می تواند، یکی از موثرترین عوامل در افزایش کیفیت امتحانات باشد.

تقدیر و تشکر

بدینوسیله از کلیه کسانی که در اجرای این تحقیق ما را یاری نمودند به ویژه اعضای محترم هیأت علمی دانشکده دندانپزشکی دانشگاه علوم پزشکی بابل تقدیر و تشکر می گردد.

[DOR: 20.1001.1.15614107.1389.12.5.10.0]

Effects of Self Assessment of the Examination Questions by Educational Groups on the Quality of Final Exams

H. Mohammadi (MD) [1*], Y. Javadian (PhD) [2], N. Nikbakhsh (MD) [3], M. Javanian (MD) [4], Z. Jahanian (MD) [1]

1. Education Development Center, Babol University of Medical Sciences, Babol, Iran
2. Department of Physiotherapy, Babol University of Medical Sciences, Babol, Iran
3. Department of Surgery, Babol University of Medical Sciences, Babol, Iran
4. Infectious Diseases and Tropical Medicine Research Center, Babol University of Medical Sciences, Babol, Iran

Received: Nov 11th 2009, Revised: Dec 9th 2009, Accepted: Mar 10th 2010.

ABSTRACT

BACKGROUND AND OBJECTIVE: Active participation of educational groups in final exam analysis has an important effect on the improvement of the question quality. This study was conducted to evaluate the effects of this self assessment on the quality of examination questions in the faculty of dentistry of Babol medical university, Babol, Iran.

METHODS: This cross sectional study was performed on 22 faculty members of dentistry faculty who attended workshop on question evaluation. They are required to evaluate the exam questions before and after each examination. Exam questions of second half of academic years of 2006-2007 and 2008-2009 which used as pre-and post test respectively were evaluated and compared.

FINDINGS: In this study 367 questions as pretest and 284 questions as post tests were evaluated. The average score for the faculty members (number of mistakes to number of questions) was 0.29±0.23 in pretest (ranging 0-0.87) and 0.18±0.14 in post tests (ranging 0-0.5). In evaluating taxonomy grade, 94% of questions of the year 2007 had taxonomy 1, 5.7% had taxonomy 2 and 0.27% had taxonomy 3. In the year 2009, 95.4% of the questions had taxonomy 1 and the others had taxonomy 2.

CONCLUSION: The results of this study showed that assessment of the questions prior to the exam by educational groups will lead to an increase in examination quality.

KEY WORDS: *Self-assessment, Examination questions, Educational groups, Final exams.*

[*] **Corresponding Author;**
Address: Department of Physiotherapy, Babol University of Medical Sciences, Babol, Iran
Tel: +98 111 2199592-6
E-mail: javad835@yahoo.com

[DOR: 20.1001.1.15614107.1389.12.5.10.0]

References

1. Zurawski RM. Making the most of exams: procedures for item analysis. The National Teaching & Learning Forum1998;7(6).http://www.ntlf.com/html/pi/9811/v7n6smpl.pdf.

2. Zolfaghari B, Adibi N, Derakhshanfar S, et al. Educational development tests in medical sciences. Isfahan, Isfahan University of Medical Sciences, Medical Education Development Management Publishing 2000. [in Persian]

3. Hooman H. Psychological and educational measurement (skill of test & questionnaire preparation), Tehran, Parsa Publishing 2002. [in Persian]

4. Gareth H, Elzubier M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Medical Education, UK, Glascow 1998; pp: 32, 343-350.

5. Haghshenas MR, Vahidshahi K, Mahmudi MN, 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 2008;5(2):119-26. [in Persian]

6. Kehoe J. Basic item analysis for multiple-choice tests. Practical assessment, research & evaluation 2007;4(10). Available at: http://pareonline.net/getvn.asp?

7. Shaban M, Ramezani Badr F. Effect of test item analysis of summative exams on quality of test designing. Hayat, J Nurse Midwifery, Tehran Univ Med Sci 2007;13(1):5-16. [in Persian]

8. Salemi L, Rafiee M. The effect of workshops of teaching method on the rate of using test design skills among teachers. Abstracts of articles in 7th Iranian Congress of Medical Education, Tabriz, Vice Chancellery for Research, Tabriz Medical University Publication 2004; pp: 54-5. [in Persian]

9. Ashraf Pour M, Beheshti Z, Molook Zadeh S. Quality of final examination in students of Babol Medical University, 1999-2000. J Babol Univ Med Sci 2003;2:42-7.[in Persian]

10. Millman J. Student achievement as a measure of teacher competence. In: Millman J (ED.), Handbook of teacher evaluation. Beverly Hills, CA: Sage Publications 1981; pp: 146-66.

11. Shakoornia A, Khosravi A, Shariati A, Zarei A. Survey on multiple choice questions of faculty members of Jondi Shapor Medical University of Ahwaz. The 8th National Congress of Medical Education. Kerman University of Medical Sciences; 2007:44. [in Persian]

12. Moattari M, Hashemi F, Parsa Yekta Z. Assessment of validity and reliability of comprehensive exam in Fatemeh Zahra nursing & midwifery faculty of Shiraz university of medical sciences, 4th Iranian Congress of Medical Education, Tehran, Vice Chancellery for Research, Tehran Medical University Publication 1999; p:61. [in Persian]

13. Javadi M, Abbas Zadeh A, Borhani F, Abdoli R. Analytical survey of summative MCQs of Bam Nursing faculty. The 8th National Congress of Medical Education. Kerman University of Medical Sciences 2007; p: 217. [in Persian]

14. Daripoor F, Karimi Sh, Pakaeen J. Evaluation of MCQs of nursing student 2006. The 8th National Congress of Medical Education. Kerman University of Medical Sciences 2007: 67. [in Persian]

15. Rasolinejad SA, Vakili Z, Fakharion E, Mosayebi Z, Moniri R. Comparative survey of taxonomies of residents. Promotion examination, Kashan Medical University 2006. The 8th National Congress of Medical Education. Kerman University of Medical Sciences 2007; p: 68.

[DOR: 20.1001.1.15614107.1389.12.5.10.0]