

M. Nafar^{*1}, M. Tayebi², L. Sadati¹, B. Pooragha³

1. Department of Surgical Technology, Instructor, School of Paramedical, Alborz University of Medical Science, Karaj, Iran

2. Department of Anesthesiology, Associate Professor, Alborz University of Medical, Karaj, Iran

3. Education Development Center, Assistant professor, Alborz University of Medical, Karaj, Iran

Assay Quantitative Indicators of Multiple-Choice Tests in Some Faculty Members of Alborz University of Medical Sciences and Its Relationship with Demographic Variables in the First Semester 95-94

Received: 3 Oct.2016; Accepted: 3 Apr.2017

Abstract

Introduction: Assessment of Educational Progress is an important part of the educational process, there are many ways to evaluate students, the most common type of Examination is multiple choice Questions and if they are properly designed, Become appropriate method for assessing learning.

Objective: This study aimed to determine quantitative indicators multiple-choice tests, some faculty members of Alborz University of Medical Sciences and its relationship with demographic variables in the first semester 95-94

Methods: In this cross-sectional study (cross sectional) of 31 faculty members of Alborz University of Medical Sciences after obtaining permission from the Vice Chancellor for Research and the University of Education so that the researchers were examined by the software descriptive and analytical tests EDC unit test analysis software based in University studied anonymous. In order to analyze the data, descriptive statistics such as mean, standard deviation, frequency distribution and Pearson and Spearman correlation coefficient was used inferential.

Results: Results showed that although the majority of tests have acceptable difficulty factor but their discrimination index is weak, the other had a small majority of test reliability.

Conclusion: The results of this study can be argued that most of the questions used in the test was in need of fundamental review and be familiar with design workshops and assessment tests in order to meet the teachers and the inevitable creation of standard questions.

Keywords: Faculty Members, Multiple-choice tests, Quantitative indicators

***Corresponding Author:**

Department of Surgical Technology,
Instructor, School of Paramedical,
Alborz University of Medical
Science, Karaj, Iran

Tel: 0912-7650117

E-mail: Nafar.mehran@yahoo.com

بررسی شاخص‌های کمی آزمون‌های چندگزینه‌ای برخی اعضای هیئت علمی دانشگاه علوم پزشکی البرز و ارتباط آن با متغیرهای دموگرافیک در نیم‌سال اول ۹۵-۹۴

تاریخ دریافت مقاله: ۹۵/۷/۱۲؛ تاریخ پذیرش: ۹۶/۱/۱۴

چکیده

مقدمه: ارزشیابی پیشرفت تحصیلی بخش مهمی از فرآیند آموزشی بوده و روش‌های مختلفی برای ارزشیابی علمی دانشجویان وجود دارد. آزمون‌های چندگزینه‌ای رایج‌ترین نوع آزمون‌ها در نظام آموزشی بوده و چنانچه به‌درستی طراحی شوند روش مناسبی برای سنجش آموخته‌های آن‌ها می‌باشند.

هدف: این مطالعه باهدف تعیین شاخص‌های کمی آزمون‌های چندگزینه‌ای برخی اعضای هیئت علمی دانشگاه علوم پزشکی البرز و ارتباط آن با متغیرهای دموگرافیک در نیم‌سال اول ۹۴-۹۵ صورت گرفته است.

روش کار: در این مطالعه توصیفی-تحلیلی (cross sectional) تعداد ۳۱ نفر از اعضای هیئت علمی دانشگاه علوم پزشکی البرز پس از کسب اجازه‌نامه از معاونت پژوهشی و معاونت آموزشی دانشگاه موردپژوهش قرار گرفتند بدین‌صورت که محقق نتایج آزمون‌های تحلیل‌شده توسط نرم‌افزار تحلیل آزمون مستقر در واحد EDC دانشگاه را به‌صورت بی‌نام موردبررسی قرارداد. به‌منظور تجزیه و تحلیل داده‌ها از آمار توصیفی مانند میانگین و انحراف معیار، جداول توزیع فراوانی و آمار استنباطی ضریب همبستگی پیرسون و اسپیرمن استفاده شد.

نتایج: یافته‌ها نشان داد که باوجود آنکه اکثر آزمون‌ها از ضریب دشواری قابل قبول برخوردار بوده اما شاخص تمیز آن‌ها در حد ضعیف می‌باشد. از سوی دیگر اکثر آزمون‌ها از ضریب پایایی کمی برخوردار بودند.

نتیجه‌گیری: با توجه به نتایج حاصل از این پژوهش می‌توان این‌گونه استدلال نمود که اکثر سؤالات مورد استفاده در آزمون‌ها نیازمند بازبینی اساسی بوده و برگزاری کارگاه‌های آشنایی با طراحی و ارزیابی آزمون به‌منظور آشنایی اساتید و ایجاد بانک سؤالات استاندارد اجتناب‌ناپذیر می‌نماید.

کلمات کلیدی: شاخص‌های کمی، آزمون‌های چندگزینه‌ای، اعضای هیئت علمی

مهران نفر^{۱*}، مهدی طیبی آراسته^۲، لیلا ساداتی^۱، بهروز پورآقا^۲

^۱گروه اتاق عمل، مربی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی البرز، کرج، ایران

^۲گروه بی‌هوشی، دانشیار، دانشگاه علوم پزشکی البرز، کرج، ایران

^۳مرکز مطالعات و تحقیقات آموزشی، استادیار، دانشگاه علوم پزشکی البرز، کرج، ایران

* نویسنده مسئول:

گروه اتاق عمل، مربی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی البرز، کرج، ایران

۰۹۱۲-۷۶۵۰۱۱۷

E-mail: nafar.mehran@yahoo.com

مقدمه

از لحاظ حساسیت کم در مقابل حدس زدن کورکورانه و هم از نظر سهولت تصحیح پاسخنامه‌ها بهترین نوع آزمون‌های عینی است.^۱ در مطالعات مختلف انجام شده در دانشگاه‌های علوم پزشکی کشور مشخص شده که ارزشیابی‌های پایان‌ترم به عمل آمده از دانشجویان عمدتاً با استفاده از سؤالات چندگزینه‌ای انجام می‌گیرد و نقص‌های قابل ملاحظه‌ای دارند. مهارت طراحی سؤال استاندارد در حال حاضر به عنوان یک مسئله قابل بحث آموزشی مطرح می‌باشد.^{۱۱} با توجه به اهمیت و حساسیت طراحی آزمون‌های استاندارد جهت قضاوت صحیح درباره قبول یا رد دانشجویان و نقشی که این موضوع در آینده علمی و حرفه‌ای آن‌ها دارد، این مطالعه بر آن است تا با بررسی میزان تطابق آزمون‌های پایان‌ترم برگزار شده برای دانشجویان دانشگاه علوم پزشکی البرز با استانداردهای طراحی آزمون، به تحلیل شاخص‌های کمی سؤالات پرداخته تا در نهایت بتواند از داده‌های به دست آمده در جهت کشف نقاط ضعف و ارتقای مهارت‌های طراحی آزمون در مدرسان استفاده نماید.

مواد و روش‌ها

پژوهش حاضر یک مطالعه مقطعی از نوع توصیفی تحلیلی است. جامعه پژوهش شامل منتخب اعضای هیئت علمی دانشگاه علوم پزشکی البرز و نمونه پژوهش شامل کلیه سؤالات چهارگزینه‌ای یکی از آزمون‌های برگزار شده توسط اعضای هیئت علمی دانشگاه علوم پزشکی البرز در نیمسال اول ۹۵-۹۴ به شیوه نمونه‌گیری در دسترس می‌باشد که در آن نتایج تحلیل آزمون ۳۱ نفر از اعضای هیئت علمی دانشگاه علوم پزشکی البرز، مورد بررسی قرار گرفته است. ابزار گردآوری اطلاعات شامل نرم‌افزار تحلیل آزمون مستقر در واحد EDC دانشگاه بوده که منبع اطلاعات دموگرافیک (جنس، سابقه تدریس و مرتبه علمی) نیز می‌باشد. داده‌های آزمون جهت ارزشیابی در نرم‌افزار تحلیل آزمون جهت استخراج ضریب دشواری، شاخص تمیز و پایایی وارد شده است. شاخص دشواری یا سهولت سؤال عبارت است از نسبت افرادی که به یک سؤال پاسخ صحیح داده‌اند.

شاخص دشواری یک شاخص وارونه است. هر چه مقدار این شاخص کوچک‌تر باشد (به صفر نزدیک‌تر باشد) سؤال مشکل‌تر و

یکی از مهم‌ترین وظایف اعضای هیئت علمی، ارزیابی علمی دانشجویان می‌باشد. این وظیفه مهم نیاز به مهارت‌های قابل توجهی دارد.^۱ ارزشیابی عبارت است از فرآیند نظام‌مند گردآوری، تحلیل و تفسیر داده‌ها به منظور ارزش‌گذاری و قضاوت که در مورد افراد، گروه‌ها و سازمان‌ها به عنوان ارزیابی و در مورد برنامه‌ها به عنوان ارزشیابی و در مورد فراگیران آموزشی به عنوان آزمون مورد اشاره قرار می‌گیرد.^{۳،۲}

ارزشیابی پیشرفت تحصیلی دانشجویان یکی از اهداف عمده دانشگاه‌ها و مراکز آموزش عالی کشور است زیرا بدین وسیله میزان یادگیری دانشجویان و در نهایت میزان دستیابی به اهداف آموزشی اندازه‌گیری می‌شود.^۴ مهم‌ترین وسیله اندازه‌گیری پیشرفت تحصیلی، آزمون است. آزمون‌هایی که عملکرد تحصیلی دانشجویان را اندازه می‌گیرند به دو نوع انشایی و عینی تقسیم می‌شوند. آزمون‌های چندگزینه‌ای که به آزمون‌های تستی شهرت دارند یکی از انواع آزمون‌های عینی هستند.^۵

خصوصیت اندازه‌گیری این است که در آن خطا وجود دارد. برای اینکه مقدار خطای آزمون‌ها کم شود دو ویژگی پایایی و روایی را باید در نظر داشت. یک آزمون هنگامی از پایایی برخوردار است که اگر آن را در موقعیت‌های متفاوت زمانی و مکانی به یک گروه از فراگیران بدهیم نتایج مشابهی به دست آید. روایی اصطلاحی است که به میزان دستیابی به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است اشاره می‌کند.^{۱۴} سؤالاتی که در اکثر دانشکده‌های علوم پزشکی مورداستفاده قرار می‌گیرند عمدتاً به روش چندگزینه‌ای طراحی می‌شوند و چنانچه با دقت طراحی شوند، می‌توانند تمام اهداف درسی را در سطوح مختلف فکری پوشش دهند.^۶ آزمون‌های چندگزینه‌ای طی بیست سال اخیر به طور چشمگیری همگانی شده است.^۷ آزمون‌های چندگزینه‌ای شامل تعدادی سؤال است که هر یک از آن‌ها از یک قسمت اصلی (تنه سؤال) و تعدادی گزینه (پاسخ) تشکیل می‌شوند و آزمون‌شونده از میان آن‌ها گزینه صحیح (پاسخ سؤال) را انتخاب می‌کند.^{۱۴} همگانی شدن آزمون‌های چندگزینه‌ای به دلیل پایایی این نوع آزمون است.^{۹،۸} این نوع آزمون هم از لحاظ یکنواختی سؤال‌ها و هم

ترکتر از آزمون‌شونده ضعیف‌تر به سؤال مزبور پاسخ درست داده‌اند یعنی سؤال قدرت تمیز ندارد و گمراه‌کننده یا مبهم است. سؤال‌های با قدرت تمیز منفی، فاقد ارزش هستند و باید حذف شوند. پایایی یک ابزار اندازه‌گیری، به درجه ثبات، همسانی و قابلیت اعتماد نتایج حاصل از آن اشاره دارد.

یک آزمون در صورتی پایایی دارد که اگر آن را در یک فاصله زمانی کوتاه چندین بار به گروه واحدی از افراد بدهیم نتایج حاصل نزدیک به هم باشند. اگر آزمونی در هر بار اجرا بر روی تعدادی آزمودنی نتایج مختلفی به دست بدهد آن آزمون یک آزمون پایا نخواهد بود و در واقع هیچ چیزی را به درستی اندازه نخواهد گرفت. در مورد پایایی سؤال نیز همین تفسیر صدق می‌کند به این معنی که اگر نمرات بدست آمده از سؤال در موارد مختلف نتایج مختلفی به همراه داشته باشد آن سؤال، سؤال پایایی نخواهد بود. حداکثر ضریب پایایی سؤال ۱ است و سؤالات با ضریب پایایی بالای ۰/۷ مطلوب است. در این مطالعه جهت سنجش پایایی سؤالات از روش آلفای کرون باخ استفاده شد.

جهت روابط سنجی میان متغیرهای دموگرافیک (جنس، مرتبه و سابقه استاد) با میزان شاخص‌های کمی آزمون از ضریب همبستگی پیرسون و اسپیرمن استفاده شده است.

نتایج

یافته‌های پژوهش نشان‌دهنده آن بود که از مجموع ۳۱ شرکت‌کننده در پژوهش ۳۸٪ شرکت‌کنندگان مرد و ۶۲٪ زن بودند. همچنین تعداد ۱۳ نفر دارای سابقه زیر ۵ سال، ۱۳ نفر بین ۵-۱۰ سال، ۳ نفر بالای ۱۰ سال و ۲ نفر بدون اطلاعات در این زمینه بودند. در ارتباط با مرتبه شغلی تعداد ۱۳ نفر مربی، ۱۴ نفر استادیار، ۲ نفر دانشیار و ۲ نفر بدون ثبت اطلاعات بودند.

میانگین نمرات شاخص دشواری نشان داد که ۶/۵٪ سؤالات آزمون‌ها دارای سطح دشواری ۰/۳-۰ (بسیار دشوار)، ۵۸/۱٪ دارای سطح دشواری ۰/۷-۰/۳ (مناسب) و ۳۵/۵٪ دارای سطح دشواری ۰/۷-۰ (بسیار آسان) می‌باشند (نمودار ۱).

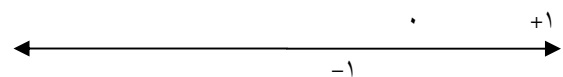
هرچه مقدار آن بزرگ‌تر باشد (به ۱۰۰ نزدیک‌تر باشد) سؤال آسان‌تر است. در آزمون‌های هنجاری سؤال‌های مناسب و قابل قبول سؤال‌هایی هستند که سطح دشواری آن‌ها در حد متوسط باشد. سؤال‌های با سطح دشواری متوسط امکان پراکندگی در نمره آزمون‌شوندگان را افزایش می‌دهند. سطح دشواری مطلوب برای یک سؤال بدون توجه به نوع سؤال حدود ۵۰ درصد (بین ۰/۳ تا ۰/۷) در نظر گرفته می‌شود.

دامنه درجه دشواری

- آسان: ۰/۷-۱
- مناسب: ۰/۷-۰/۳
- دشوار: ۰/۳-۰

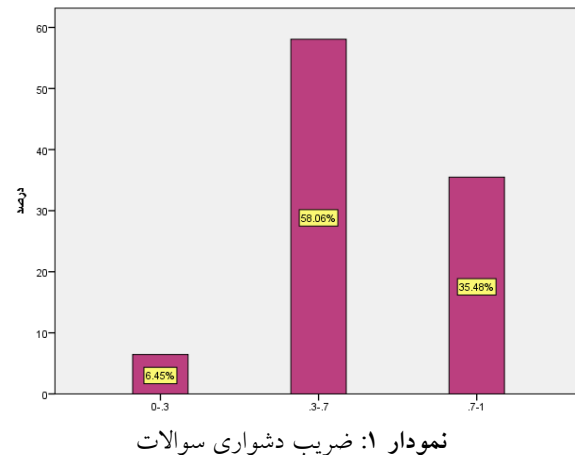
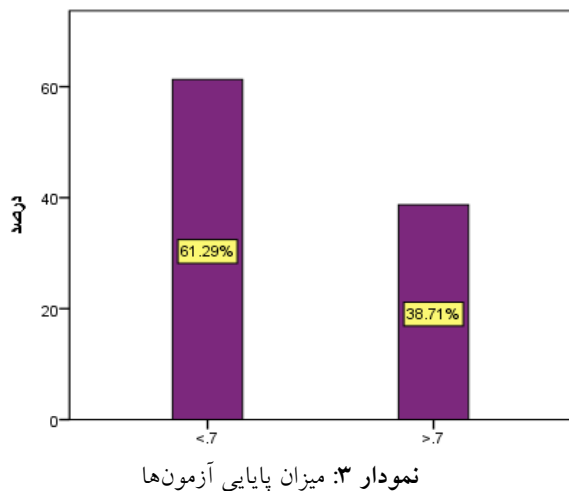
بهترین سطح دشواری بین ۰/۷-۰/۳ در نظر گرفته خواهد شد، زیرا حداکثر اطلاع را درباره تفاوت بین آزمون‌شوندگان به دست می‌دهد. قدرت تمیز یا تفکیک سؤال عبارت است از میزان قدرت سؤال در تفکیک آزمون‌شوندگان دارای نمره بالا از آزمون‌شوندگان دارای نمره پایین. به عبارت دیگر این شاخص مشخص می‌نماید که آزمون چقدر توانسته است بین گروه قوی و ضعیف افتراق بگذارد. هرچه عدد ضریب تمیز بزرگ‌تر باشد سؤال از توان بیشتری در جدا کردن و تفاوت قائل شدن میان گروه قوی و ضعیف برخوردار می‌باشد. منفی بودن ضریب تمیز به معنی آن است که گروه قوی از گروه ضعیف عملکرد پایین‌تری داشته‌اند. این مشکل به‌طور معمول ناشی از ایراد خود سؤال می‌باشد و چنانچه سؤال دارای اشکال فنی نباشد، افراد گروه بالا یا آن سؤال را به‌طور کامل یاد نگرفته‌اند یا آن را به‌غلط آموخته‌اند.

دامنه قدرت تمیز سؤال بر اساس نمودار ذیل تفسیر می‌شود:



- مبهم و گمراه‌کننده: ۰...-۱
- خیلی ضعیف: ۰/۲-۰
- ضعیف: ۰/۴-۰/۲
- مناسب: ۰/۴-۱

دامنه تغییر ضریب تمیز از +۱ تا -۱ می‌باشد که اعداد مثبت (+) مطلوب ماست. اعداد منفی (-) نشان می‌دهد آزمون‌شونده قوی



نتایج به‌دست‌آمده از آزمون ضریب همبستگی اسپیرمن نشان داد که از بین متغیرهای مرتبه و جنس، تنها مرتبه اساتید با سطح دشواری سوالات دارای ارتباط معنادار آماری بود ($r = -0/508$ و $p = 0/005$) به نحوی که با افزایش مرتبه اساتید سطح دشواری سوالات افزایش می‌یابد.

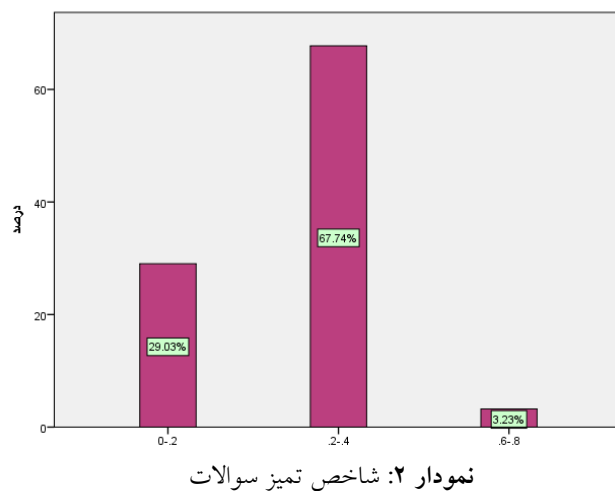
نتایج شاخص تمیز سوالات گویای آن بود که ۲۹٪ سوالات دارای شاخص تمیز ۰-۰/۲ (خیلی ضعیف)، ۶۷/۷٪ دارای شاخص تمیز ۰/۲-۰/۴ (ضعیف) و ۳/۲٪ دارای شاخص تمیز ۰/۴-۱ (مناسب) می‌باشد (نمودار ۲).

در ارتباط با شاخص پایایی، ۶۱/۳٪ آزمون‌ها دارای نمره پایایی کمتر از ۰/۷ و ۳۸/۷٪ نمره بالای ۰/۷ داشتند ($0/57 \pm 0/22$ میانگین) (نمودار ۳).

بحث و نتیجه‌گیری

از داده‌های به‌دست‌آمده چنین استنباط می‌شود که با وجود اینکه بیش از نیمی از سوالات از سطح دشواری قابل قبول (۰/۵۸/۱) برخوردارند اما اکثریت آن‌ها (۶۷/۷٪) دارای شاخص تمیز ضعیف (۰/۲-۰/۴) می‌باشند که این امر نشان‌دهنده قدرت کم آزمون‌ها در تمایز دانشجویان ضعیف و قوی از یکدیگر می‌باشد. در مطالعه تشنیزی و همکاران که با عنوان تجزیه و تحلیل کیفیت سوالات آزمون‌های چندگزینه‌ای رشته کارشناسی ناپیوسته مدارک پزشکی در سال ۸۹ بر روی تمامی آزمون‌های چندگزینه‌ای پایان‌ترم مربوط به رشته مدارک پزشکی دانشگاه علوم پزشکی هرمزگان انجام شد نتایج حاکی از آن بود که هرچند ضریب دشواری برای تمام سوالات آزمون‌ها از میزان قابل قبول و مناسب برخوردار است اما شاخص تمیز تقریباً تمامی سوالات آزمون‌ها در حد ضعیف تا خیلی ضعیف بود. محقق این‌گونه نتیجه‌گیری می‌کند که بهتر است این سوالات از بانک سوالات حذف‌شده یا بطور اساسی مورد تجدیدنظر قرار گیرد.^۳

نتایج به‌دست‌آمده از ضریب همبستگی پیرسون هیچ رابطه معناداری را میان سابقه اساتید با شاخص‌های تحلیل آزمون (دشواری، تمیز و پایایی) نشان نداد.



افزایش مرتبه استاد بر دشواری سؤالات افزوده می‌شود که این امر احتمالاً به علت رویکرد تخصصی تر ایشان نسبت به مسئله طرح سؤالات آزمون می‌باشد.

به عقیده پژوهشگر برگزاری مستمر دوره‌های آموزشی طراحی آزمون، ایجاد بانک سؤالات استاندارد، بازخورد نتایج تحلیل آزمون به اساتید و لحاظ نمودن ارزیابی آزمون‌ها در نمرات ارزشیابی اساتید می‌تواند نقش بسزایی در افزایش کیفیت آزمون‌های چهارگزینه‌ای و به تبع ارزیابی بهتر دانشجویان داشته باشد.

تشکر و قدردانی

این مقاله حاصل طرح تحقیقاتی مصوب دانشگاه علوم پزشکی و خدمات بهداشتی و درمانی البرز (شماره ثبت ۲۷۱۸۵۹۹) می‌باشد. بدین وسیله محققین مراتب تشکر و قدردانی خود را از معاونت محترم پژوهش و فناوری دانشگاه علوم پزشکی گیلان به دلیل تصویب و تأمین هزینه اجرای طرح اعلام داشته و از اعضای هیئت علمی دانشگاه بابت شرکت در پژوهش سپاسگزاری می‌نمایند.

با توجه به سطح پایایی بالای ۰/۷ به عنوان سطح قابل قبول می‌توان نتیجه گرفت که اکثر آزمون‌ها از نمره پایایی قابل قبول برخوردار نمی‌باشند که این مورد می‌تواند به علت عدم تجانس آزمون‌شوندگان (تفاوت در توانایی پاسخ‌دهی به سؤالات)، تعداد کم گزینه‌های سؤال و یا تعداد کم سؤالات باشد که نیازمند بررسی مجدد در این زمینه‌ها می‌باشد. در پژوهش یراقی و همکاران با عنوان بررسی شاخص‌های کمی و کیفی آزمون‌های مقطع کارآموزی دوره پزشکی عمومی در سال ۸۷ میانگین ضریب پایایی ۷۷/۹۲ درصد بود. نکته قابل توجه برای اساتید دانستن این نکته است که با انتخاب سوالاتی با ضریب تمیز بالا و ضریب دشواری متوسط، ضریب پایایی آزمون افزایش پیدا می‌کند.

هرچقدر ضریب پایایی آزمون به ۱ نزدیک‌تر باشد از میزان واریانس خطا کاسته و نشان‌دهنده این است که واریانس مشاهده‌شده واریانس حقیقی است نکته قابل توجه در ارتباط سنجی میان متغیرهای دموگرافیک و شاخص‌های آزمون، وجود ارتباط معنادار میان مرتبه استاد (مربی، استادیار و دانشیار) با دشواری سطح سؤالات می‌باشد به نحوی که با

منابع

1. of summative exams on quality of test designing. The Journal of Faculty of Nursing & Midwifery. 2007;13:5-15
2. Seif A. Educational measurement, assessment and evaluation. 4th ed Tehran: Doran; 2004: 426 – 8 [In Persian]
3. Bazargan A. Educational assessment. Tehran: SAMT; 2001: 150-3 [In Persian]
4. Hosseini T, Zare S, Solati Seyed M. Analyzing the quality of the questions of the Bachelor of Medical Records multi-choice tests. Hormozgan Medical Journal. Fourteenth, Number 3 (Autumn 2013): 177-183. [In Persian]
5. Greg Yaraghi M, Avicenna M, Ebrahimi A, Esfandiari E, Ismaili A, Shayan Sh and others. Evaluation of quantitative and qualitative indices of general medical practice apprenticeship. Journal of Education in Medical Sciences, 2010; 5: 533-542.
6. Schultheis NH. writing cognitive educational objectives and multi choice tests questions. Am J health system Pharm. 1998;55:2397-2401
7. Zolfaghari B et al. Academic achievement tests in medical sciences, Isfahan: Isfahan University of Medical Sciences, Isfahan, Iran; 1390. 59-78. [In Persian]
8. Falahati M, Mohammadzadeh Z, Entezari M H. The quality and credibility of conducting quizzes at the examinations. Journal of Education in Spring Medicine, Eighth Year, 2005: 199-199.
9. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach 2010 Dec;26(8):709-12.
10. Azizi F, Medical Education, Challenges and Perspectives, Tehran: Deputy of Education and Student Affairs Ministry of Health and Medical Education; 1393. 686-96. [In Persian]
11. Salemi A, Rafiei M. Effects of workshops on teaching methods of using skill of design questions among teachers. Special proceedings of the seventh national conference on medical education: 2004; Tehran, Iran.
12. Yousefi M, Ahari A. Marv comprehensive biostatistics. Tehran: Sectarian Publishing; First Edition, Second Edition, 1392. 10-10. [In Persian]

13. Shakurnia A H, Mozafari A, Khosravi B A. Investigating the structure of multiple choice questions in the Assistant Exam in Jundishapur University of Medical Sciences, Ahvaz. *Journal of Medical Sciences*, 2012; 8(4): 491-592.
14. Gholami A, Derakhshan F, Safdari F. Principles for designing and evaluating multiple choice questions (MCQ). Qazvin University of Medical Sciences Publications. Summer 2011, p. 20-9.[In Persian]
15. Haghshenas MR, Vahid Shahi C, Mahmoudi M, Shahbaz Nejad L, et al. Multivariate tests of Mazandaran University of Medical Sciences in the first half of the academic year 86-85. Developmental steps in medical education, *Journal of Medical Education Development Center*, 2009; 5(2):116-26.
16. Imam Khomeini M, Zahedifar F. Analyzing the Multiple Choice Questions of Health Sciences at the Faculty of Health of Qazvin University of Medical Sciences in Qazvin, 2009. *Journal of Qazvin University of Medical Sciences*, 2010, No. 4, Winter 2011: 110-109.[In Persian]