مجله مرکز مطالعات و توسعه آموزش علوم پزشکی یزد
دورهٔ هشتم، شماره ۴، شماره پیاپی ۱۲، زمستان ۹۲
ص ۷۵-۶۴

بررسی روند پنج ساله طراحی آزمون‌های چندگزینه‌ای ارتقاء دستیاری تخصصی پزشکی دانشگاه علوم پزشکی کرمانشاه از نظر کمی و کیفی: ۱۳۹۱-۱۳۸۷

رضا پورمیرزا کلهری[۱]، دکتر منصور رضایی[۲]، بهزاد کرمی متین[۳] ، فرح روشن پور[۴*]

۱- کارشناس ارشد پرستاری، گروه فوریت‌های پزشکی، عضو هیئت علمی دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
۲- دکترای آمار زیستی، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
۳-دکترای مدیریت خدمات بهداشتی و درمانی، گروه بهداشت عمومی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
۴- کارشناس ارشد تغذیه، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

تاریخ دریافت: ۱۳۹۲/۵/۲۷ تاریخ پذیرش: ۱۳۹۲/۹/۱۸

چکیده

سابقه و اهداف: یکی از روش‌های ارزشیابی در ارتقاء دستیاری پزشکی آزمون‌های چند گزینه‌ای است. این مطالعه با هدف بررسی روند پنج ساله طراحی آزمون‌های چند گزینه‌ای ارتقاء دستیاری تخصصی پزشکی دانشگاه علوم پزشکی کرمانشاه از نظر کمی و کیفی (در سال‌های ۹۱-۱۳۸۷) انجام شد.

روش بررسی: مطالعه از نوع توصیفی- تحلیلی -طولی از نوع روند پژوهی بود. ۶۳ آزمون (۹۶۰۰ سؤال) از لحاظ شاخص‌های ضریب دشواری، ضریب تمییز و اعتبار کل آزمون، درصد تاکسونومی I و II و III و درصد سؤالات بدون اشکال ساختاری بررسی شدند. ابزار جمع آوری داده‌ها خروجی نرم افزار آنالیز کمی و گزارشات آنالیز کیفی بودند که با نرم افزار SPSS V18 و آزمون‌های t- test، کای دو، ANOVA و کروسکال والیس تحلیل شدند.

یافته‌ها: در طول پنج سال میانگین‌های ضریب دشواری کل سؤالات ۰/۶۲، ضریب تمییز ۰/۲۷ و ضریب اعتبار ۰/۸۸ و در محدوده قابل قبول بود، ولی میانگین‌های درصد تاکسونومی I سؤالات ۳۳/۴۴٪، تاکسونومی II و III ۶۶/۵۵٪ و درصد سؤالات بدون اشکال ساختاری ۶۲/۵۹٪ بود که کمتر از محدوده قابل قبول بود. همچنین تفاوت آماری معنی داری بین شاخص‌های ضریب دشواری (P=۰/۰۰۱)، اعتبار کل آزمون (P=۰/۰۱۷)، میانگین تاکسونومی I (P=۰/۰۰۱) و تاکسونومی II و III (P=۰/۰۰۶) در طول پنج دوره آزمون وجود داشت.

نتیجه‌گیری: بر اساس یافته‌های این مطالعه، کاهش میانگین تاکسونومی II و III و افزایش تاکسونومی I در دوره پنج ساله بررسی آزمون‌های ارتقاء دستیاری پیشنهاد کننده استفاده از برنامه‌های آموزشی ارتقاء دانش علمی طراحان آزمون‌های ارتقاء دستیاری است.

واژه‌های کلیدی: سؤالات چندگزینه‌ای، آنالیز آزمون، دستیاری پزشکی

* (نویسنده مسئول)؛ تلفن:۰۹۱۸۸۳۹۳۷۳۲، آدرس الکترونیکی: froshanpour7@ymail.com

ارجاع به این مقاله به صورت زیراست:
Pourmirza kalhori R, Rezaei M, Karami Matin B, Roshanpour F. ***A survey of quality and quantity indexes of multiple choice question (MCQ) exams of medical residents at Kermanshah University of Medical Sciences: 2008-2012.*** Journal of Medical Education and Development. 2014; 8(4): 64-75

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

مقدمه

هدف نهایی آموزش ایجاد یادگیری پایدار در فراگیر است و ابزار سنجش یادگیری، آزمون است (۱،۲). هر آزمون دارای سه شاخص روایی، پایایی و عملی بودن است (۳) و در دو دسته کلی آزمون‌های معلم ساخته و آزمون‌های استاندارد طبقه‌بندی می‌شود (۴،۵). آزمون‌های چندگزینه‌ای متداول‌ترین آزمون‌های عینی در علوم پزشکی هستند (۳). آزمون‌های عینی اگر توسط افراد مختلف و یا در زمان‌های مختلف تصحیح شود دارای نتایج یکسانی هستند. هر سؤال چندگزینه‌ای دارای بخش ریشه و گزینه‌های پیشنهادی است (۲). از سؤالات چندگزینه‌ای برای ارزشیابی معلومات فردی، تصمیم گیرهای پزشکی، تفسیرهای آماری و مهارت‌های ذهنی مانند یادآوری، تشخیص و مشکل‌گشایی استفاده می‌شود (۵) و اگر سؤالات به خوبی طراحی شوند می‌توانند سطوح دانش، فهم، ادراک، کاربرد و اطلاعات را نیز ارزشیابی کنند (۶). رایج‌ترین نوع آزمون کتبی در آموزش پزشکی آزمون‌های چندگزینه‌ای است که از حدود پنج دهه قبل در آزمون کتبی بوردهای تخصصی ایالات متحده مورد استفاده قرار می‌گیرند (۷،۸).در ایران نیز در آموزش پزشکی آزمون‌های چندگزینه‌ای متداول‌ترین آزمون‌های کتبی در ارزشیابی دانشجویان گروه علوم پزشکی است(۵).

در رشته‌های علوم پزشکی که به طور مستقیم با سطح سلامت مردم و جامعه در ارتباط هستند، ارتقاء کارایی و سطح علمی و عملی فارغ‌التحصیلان هدف نهایی آموزش است و کیفیت آزمون‌ها در دستیابی به این هدف دارای تأثیر فراوانی است به طوری که راید و همکاران (Reid et al)، (۲۰۰۷) و لارسن و همکاران (Larsen et al)، (۲۰۰۸) گزارش کرده‌اند که کیفیت آزمون‌ها، بر کیفیت آموزش و روند یاددهی – یادگیری موثر است (۱۱-۹).

بعد از برگزاری آزمون‌های چندگزینه‌ای، تجزیه و تحلیل سؤالات آزمون از نظر شاخص‌های کمی و کیفی برای تعیین مناسبت سؤالات و تشکیل بانک سؤالات، به خصوص در آزمون‌های ملاک نسبی (وابسته به هنجار) مانند آزمون‌های ارتقاء دستیاری مورد تأکید فراوان است (۱۲).

تحلیل کمی آزمون‌ها با محاسبه ضریب پایایی، ضریب دشواری و ضریب تمیز انجام می‌شود. تحلیل کیفی آزمون شامل درصد تاکسونومی I و II و III سؤالات و درصد سؤالات بدون اشکال ساختاری است. جهت طراحی سؤالات چند گزینه بدون اشکال ساختاری باید اصول میلمن رعایت شود. مهم‌ترین اصول میلمن در طراحی سؤالات چندگزینه‌ای عبارتند از: درج قسمت اصلی اطلاعات سؤال در ساقه سؤال، پیگیری هر هدف یادگیری با یک سؤال، استفاده از لغات ساده و روشن، عدم کاربرد گزینه منفی در ساقه سؤال، عدم کاربرد گزینه همه موارد یا هیچ‌کدام، عدم کاربرد گزینه متضاد، مشخص کردن کلمات منفی، استقلال سؤالات از یکدیگر، همسانی سؤالات از لحاظ طول و ساختار لغوی، عدم کاربرد کلمات تکراری، نداشتن غلط املایی و درج گزینه‌ها به صورت عمودی می‌باشند. درصد تاکسونومی سؤالات منطبق بر اهداف آموزشی شامل: تاکسونومی I (سطح دانش بکار رفته)، تاکسونومی II (یادآوری) و تاکسونومی III (کاربرد مفاهیم) است (۴).

در ایران مطالعات مختلفی در علوم پزشکی به تجزیه و تحلیل آزمون‌های چندگزینه‌ای مدرس ساخته پرداخته‌اند که می‌توان به گزارش حداقل ضریب پایایی و ضریب‌های دشواری در آزمون‌های چندگزینه‌ای کارآموزی‌های بالینی پزشکی (۱۳) و بالا بودن ضریب دشواری و مناسبت اعتبار سؤالات آزمون‌های فیزیوپاتولوژی اشاره کرد (۱۴).

تجزیه و تحلیل ۱۵۸۶۰ سؤال آزمون‌های پایان نیمسال ۱۳۸۷-۱۳۸۶ دانشکده پزشکی قزوین نشان داد که ۱۶/۳ درصد سؤالات از نظر ضریب دشواری و تمیز مناسب بودند (۱۵) و در آنالیز ۲۴ آزمون دانشکده پزشکی مشهد، ۵۲/۲ درصد سؤالات ضریب دشواری نامناسب و ۴۹/۲ درصد موارد ضریب تمیز غیرقابل‌قبول داشتند (۱۶). نامناسب بودن ضریب‌های دشواری و تمیز سؤالات چندگزینه‌ای طراحی‌شده

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

در دانشکده‌های پزشکی می‌تواند ناشی از اشکالاتی در زمینه طراحی سؤالات چندگزینه‌ای باشد (۱۷).

در مورد کیفیت سؤالات نیز شرایط چندان ایده آل نیست. در دانشکده پزشکی مازندران گزارش شده است که ۴۶٪ از سؤالات چندگزینه‌ای بدون اشکال ساختاری هستند (۱۸) و بیشترین مشکلات ساختاری در آنالیز کیفی برخی سؤالات دانشگاه کرمانشاه تکراری بودن کلمات در تنه سؤال و مشخص نکردن عبارات منفی گزارش شده است(۱۹). در سایر کشورها نیز گزارشات متعددی از آنالیز کمی و کیفی سؤالات چند گزینه‌ای وجود دارد.

در دانشکده پزشکی کوالالامپور مالزی یک سوم از سؤالات دارای ضریب دشواری ۰/۸ بودند که نشان‌دهنده آسان بودن سؤالات برای دانشجویان است (۲۰). در آرژانتین سؤالات چندگزینه‌ای آموزش برونکوسکوپی دارای میانگین ضریب دشواری ۰/۶۵ و میانگین ضریب تمیز ۰/۵۲ و پایایی ۰/۷۵ بود که نشانه قابل‌قبول بودن آزمون‌ها در این شاخص‌ها است (۲۱).

در زمینه آنالیز سؤالات آزمون‌های ارتقاء دستیاری کشور ایران برخی شاخص‌های کیفی سؤالات چندگزینه‌ای آزمون کتبی ۲۵ رشته تخصصی طی سال‌های ۸۸-۸۶ بررسی شدند و تداوم و گسترش برنامه ارتقاء کیفیت آزمون‌ها جهت بهبود کیفیت ساختاری و ضریب تمایز سؤالات مورد تأکید قرار گرفته است (۲۲) و تأثیر مثبت آموزش اعضاء هیئت علمی در طراحی آزمون‌های ارتقاء دستیاری (۲۳) در برخی مطالعات گزارش شده است. در خارج از کشور نیز بررسی‌های سؤالات آزمون‌های دستیاری رادیولوژی (۲۴)، بیهوشی (۲۵) و سایر گروه‌های آموزش پزشکی (۲۶) منعکس‌کننده وجود اشکالات ساختاری در طراحی سؤالات و عدم پوشش اهداف متعالی آموزش است.

با تمرکز بر بررسی روند طراحی آزمون‌های ارتقاء دستیاری در دانشگاه‌های علوم پزشکی کشور و خاطرنشان کردن این نکته که هر دانشگاه به طور مستقل اقدام به برگزاری و طراحی آزمون‌ها ارتقاء دستیاری در گروه‌های آموزشی دارای دستیار توسط اعضاء هیئت علمی می‌نماید (بدین صورت که ممکن است چندین سال متوالی طراح سؤالات آزمون ارتقاء در گروه آموزشی یکسان باشند) این مطالعه با هدف کلی تعیین روند پنج ساله طراحی آزمون‌های چندگزینه‌ای ارتقاء دستیاری تخصصی پزشکی دانشگاه علوم پزشکی کرمانشاه از نظر کمی و کیفی در طول سال‌های ۱۳۹۱-۱۳۸۷ انجام شد.

روش بررسی

این مطالعه از نوع توصیفی- تحلیلی است که به روش روند پژوهی انجام شد. جامعه مورد مطالعه، مجموع کل سؤالات آزمون‌های ارتقاء دستیاری تخصصی رشته‌های پزشکی دانشگاه علوم پزشکی کرمانشاه بود که توسط اعضای هیئت علمی دانشگاه علوم پزشکی کرمانشاه در طول پنج دوره از سال ۱۳۸۷ تا ۱۳۹۱ طراحی‌شده بودند. روش جمع‌آوری نمونه‌ها به صورت سرشماری بود و بر اساس اطلاعات موجود از ۱۲ رشته تخصصی پزشکی (داخلی، جراحی، بیهوشی، زنان و زایمان، قلب، ارولوژی، عفونی، رادیولوژی، نورولوژی، آسیب‌شناسی، روان‌پزشکی و رادیولوژی) ۶۳ آزمون مورد بررسی قرار گرفت. این تعداد آزمون در گروه‌های دستیاری دانشگاه علوم پزشکی کرمانشاه برگزار و در واقع کل آزمون‌های موجود در طول پنج سال بود. تعداد ۹۶۰۰ سؤال در مجموع مورد آنالیز قرار گرفتند.

برای جمع‌آوری داده‌های کمی از گزارش‌های نرم‌افزار آنالیز کمی آزمون‌های ارتقاء دانشکده پزشکی استفاده شد. این برنامه نرم‌افزاری آزمون‌های وابسته به هنجار را با استفاده از نظریه کلاسیک آزمون‌ها Classical Test Theory (CTT) تحت تحلیل کمی قرار می‌دهد که در تمامی دانشکده‌های پزشکی کشور برای آنالیز کمی سؤالات آزمون‌های دستیاری مورد استفاده قرار می‌گیرد و مورد تأیید دبیرخانه تخصصی بررسی آزمون‌های تخصصی وزارت بهداشت درمان و آموزش پزشکی است (۲۲).

داده‌های کیفی با استفاده از گزارش‌های واحد ارزشیابی مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

کرمانشاه جمع‌آوری شدند. گزارش‌ها حاوی تحلیل کیفی آزمون‌های ارتقاء دستیاری برای ارسال به دبیرخانه تخصصی آزمون‌های ارتقاء دستیاری وزارت بهداشت درمان و آموزش پزشکی است.

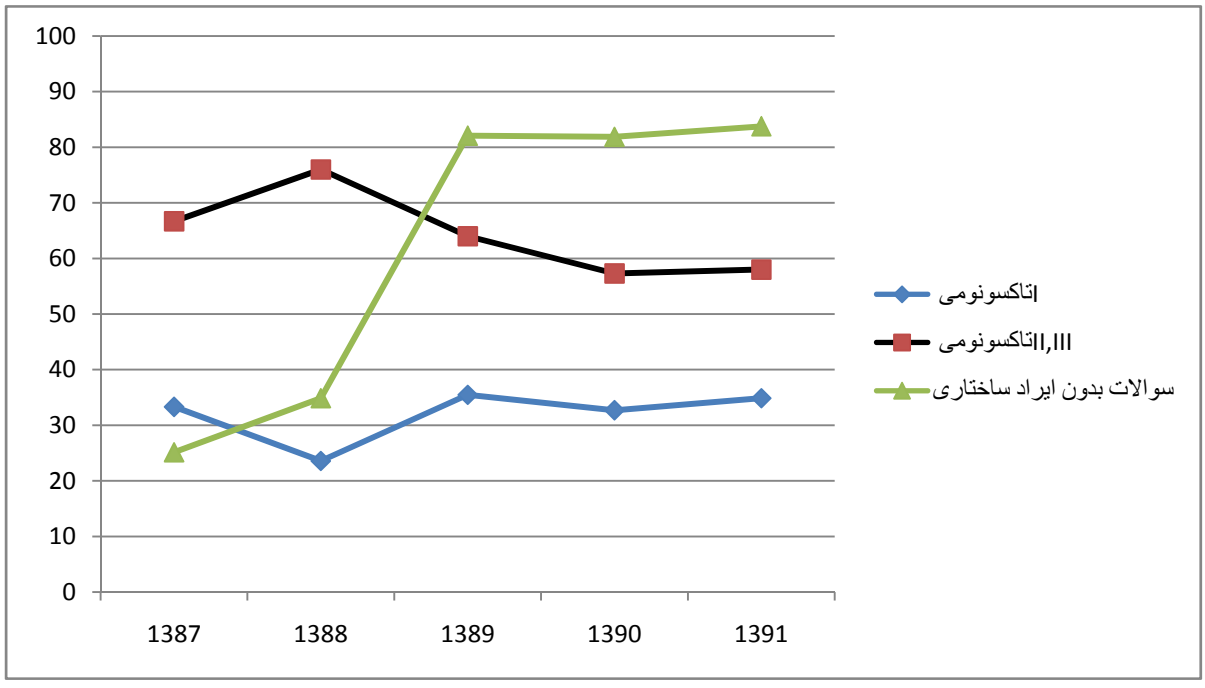
داده‌ها با استفاده از نرم‌افزار آماری SPSS V18 تحلیل شدند و برای گزارش بهتر داده‌ها از میانگین و انحراف معیار برای شاخص‌های کمی و درصدها برای متغیرهای کیفی استفاده شد. برای تحلیل سؤالات بر حسب ویژگی‌های کمی و کیفی، جداول دو بعدی به صورت تعداد و درصد و برای مقایسه با استاندارد از آزمون مجذور کای، آنالیز واریانس و کروسکال والیس استفاده شد.

در این مطالعه برای رعایت ملاحظات اخلاقی و حفظ مقام مدرسین به اسامی مدرسین طراح سؤالات اشاره نشد و مقایسه گروه‌ها نیز با یکدیگر نیز انجام نشد. برای انجام این پژوهش از ریاست مرکز مطالعات و توسعه آموزش علوم پزشکی دانشگاه علوم پزشکی کرمانشاه مجوزهای لازم اخذ شد.

یافته‌ها

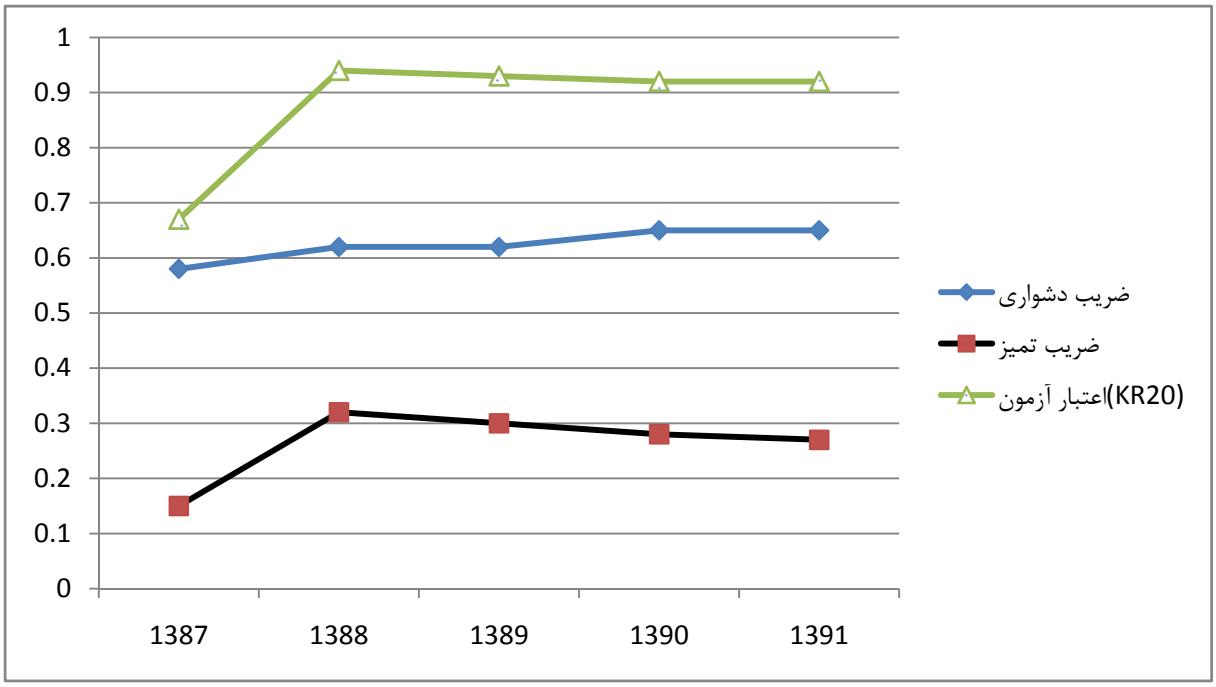
شاخص‌های کمی کل آزمون‌های بررسی‌شده در گروه‌های آموزشی نشان داد که از لحاظ ضریب دشواری ۵۴ آزمون (۸۵/۷٪) در محدوده قابل‌قبول ۰/۷-۰/۳ و از لحاظ ضریب تمیز ۴۹ آزمون (۷۷/۸٪) در محدوده ۰/۴-۰/۲ بودند و ۵ آزمون (۷/۹٪) ضریب بالاتر از ۰/۴ داشتند و از نظر اعتبار کل آزمون (ضریب KR20) ۶۲ آزمون (۹۸/۴٪) اعتبار بالاتر از ۰/۵ داشتند.

از نظر شاخص‌های کیفی در کل آزمون‌ها فقط ۸ آزمون (۱۲/۷٪) دارای تاکسونومی I کمتر از ۲۰٪ و ۵۷ آزمون (۹۰/۵٪) دارای مجموع تاکسونومی II و III بیشتر از ۸۰٪ بودند. در ۳۷ آزمون (۵۸/۷٪) بیشتر از ۸۰٪ سؤالات بدون اشکال ساختاری بود. نمودارهای ۲ و ۱ میانگین شاخص‌های کمی و کیفی را در طول روند پنج ساله بررسی سؤالات آزمون‌های ارتقاء دستیاری تخصصی پزشکی در دانشگاه علوم پزشکی کرمانشاه نشان می‌دهند.



نمودار ۱: میانگین ضریب دشواری، ضریب تمیز (افتراق) و اعتبار آزمون‌های ارتقاء دستیاری دانشگاه علوم پزشکی کرمانشاه به تفکیک سال

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]



نمودار ۲: میانگین تاکسونومی I، میانگین تاکسونومی II و III و سؤالات بدون ایراد ساختاری آزمون‌های ارتقاء دستیاری دانشگاه علوم پزشکی کرمانشاه به تفکیک سال

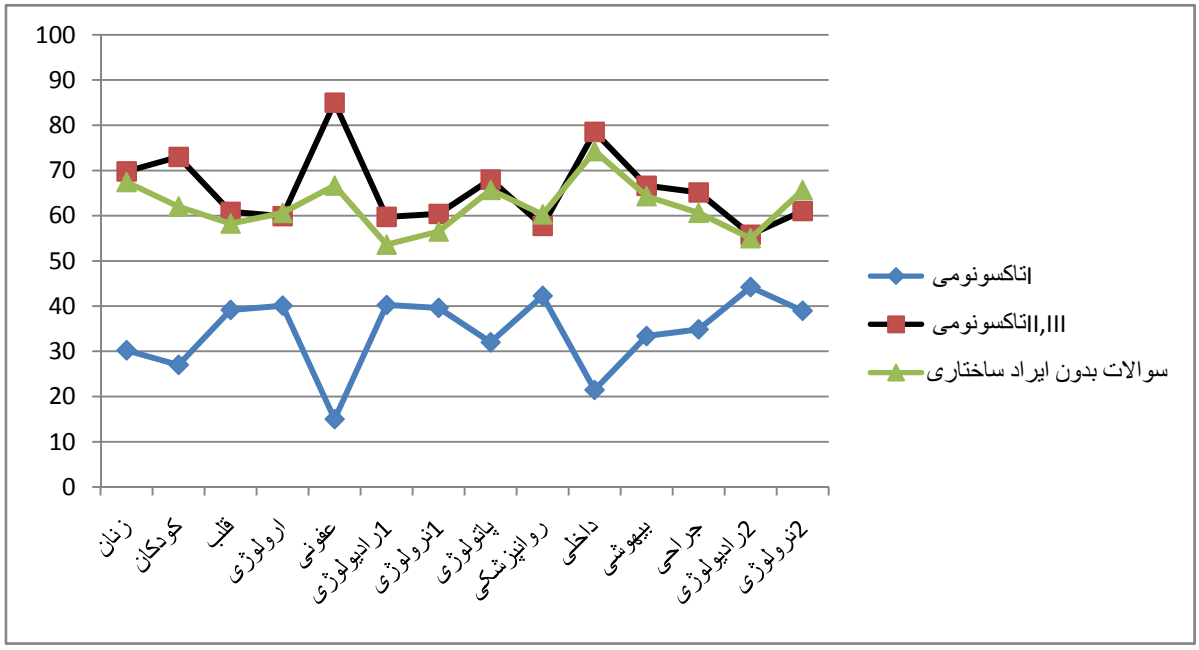
در طول پنج دوره آزمون ارتقاء دستیاری برگزار شده به تفکیک گروه‌ها و سال برگزاری آزمون، میانگین ضریب دشواری ۰/۶۲±۰/۰۶۴، میانگین ضریب تمیز ۰/۲۷±۰/۰۷۳، میانگین ضریب اعتبار ۰/۸۸±۰/۱۳، میانگین درصد سؤالات دارای تاکسونومی I ، ۳۳/۴۴±۱۲/۷۴، تاکسونومی II و III، ۶۶/۵۵±۱۲/۷۴ و میانگین درصد سؤالات بدون اشکال ساختاری ۶۲/۵۹±۲۹/۴۸ بود. جداول ۱و ۲ فاصله اطمینان ۹۵ درصدی شاخص‌های کمی و کیفی را در آزمون‌های ارتقاء دستیاری به تفکیک سال برگزاری آزمون‌ها نشان می‌دهند.

آزمون آماری آنالیز واریانس نشان داد که تفاوت آماری معنی‌داری بین شاخص کمی ضریب دشواری در طول پنج دوره آزمون برگزار شده وجود دارد (P=۰/۰۰۱). آزمون آماری کرکسال والیس این تفاوت آماری معنی‌دار را در شاخص کمی اعتبار کل آزمون‌ها را نیز نشان داد (P=۰/۰۱۷). آزمون آماری آنالیز واریانس نیز تفاوت آماری معنی‌دار در شاخص‌های کیفی میانگین تاکسونومی I (P=۰/۰۰۱) به صورت افزایش درصد تاکسونومی I سؤالات و میانگین تاکسونومی II و III (P=۰/۰۰۶) به صورت کاهش درصد این شاخص‌ها را نشان داد. تفاوت درصد سؤالات بدون اشکال ساختاری در گروه‌ها مشهود اما از لحاظ آماری معنی‌دار نبود.

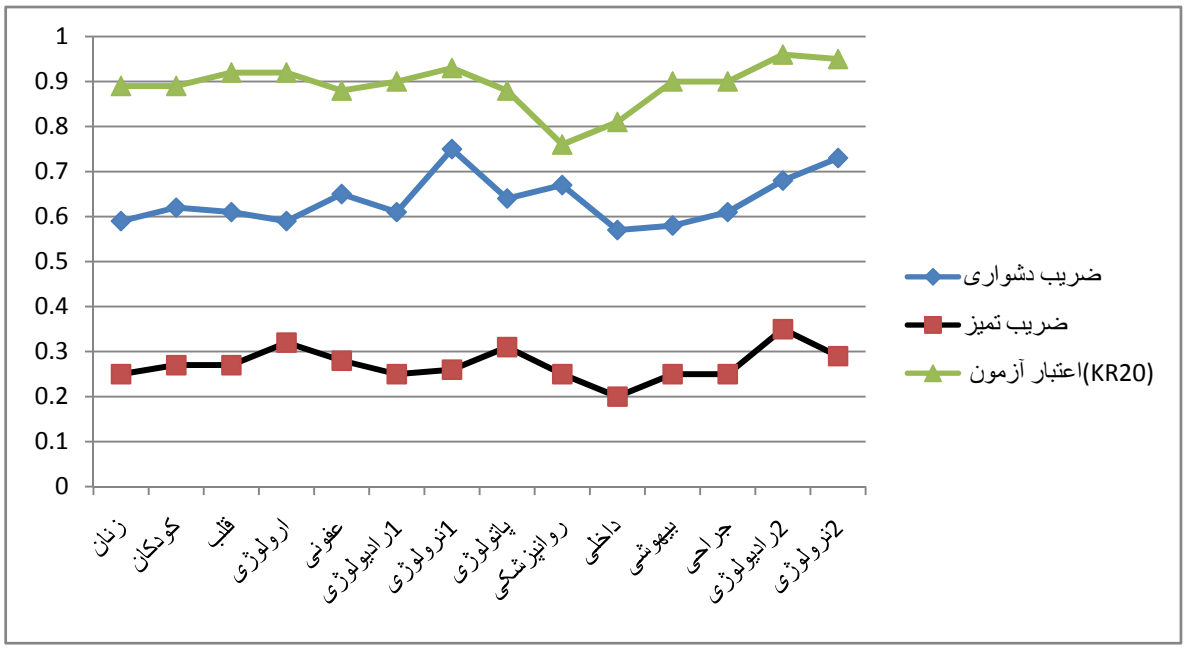
با توجه به رسالت مقاله در مورد عدم انجام مقایسه در بین گروه‌های آموزشی آنالیز استنباطی در گروه‌ها انجام نشد. اما اطلاعات دوره پنج ساله آزمون‌ها در گروه‌های آموزشی و مقادیر توصیفی شاخص‌ها در گروه ذکر می‌شوند، (نمودارهای ۳ الی ۴). سخت‌ترین آزمون‌ها در گروه داخلی و آسان‌ترین آزمون در گروه نورولوژی طراحی شده بود. بالاترین ضریب تمیز در آزمون‌های گروه رادیولوژی ۲، کم‌ترین درصد تاکسونومی I سؤالات در گروه عفونی و بیشترین در گروه رادیولوژی طراحی شده بود. کم‌ترین درصد تاکسونومی II و III سؤالات در گروه رادیولوژی ۲ و بیشترین در گروه عفونی

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

طراحی شده بود. بالاترین ضریب اعتبار آزمون در گروه رادیولوژی و بیشترین درصد سؤالات بدون اشکال ساختاری در گروه داخلی طراحی شده بود. با توجه به اهمیت رعایت استانداردهای شاخص‌های کمی و کیفی آزمون‌های ارتقاء دستیاری و لزوم بررسی این شاخص‌ها، فاصله اطمینان شاخص‌های کمی و کیفی آزمون‌های پنج ساله دستیاری بررسی‌شده در این مطالعه در جدول ۱ و ۲ ذکر شده‌اند.



نمودار ۳: میانگین تاکسونومی I، میانگین تاکسونومی II و III و سؤالات بدون ایراد ساختاری آزمون‌های ارتقاء دستیاری کرمانشاه در دوره بررسی پنج ساله به تفکیک گروه آموزشی



نمودار ۴: میانگین ضریب دشواری، ضریب تمیز و اعتبار آزمون (KR20) سؤالات آزمون‌های ارتقاء دستیاری کرمانشاه در دوره بررسی پنج ساله به تفکیک گروه آموزشی

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

جدول ۱: فاصله اطمینان ۹۵٪ برای شاخص‌های کمی آزمون‌های ارتقاء دستیاری دانشگاه علوم پزشکی کرمانشاه به تفکیک سال

دانشکده / شاخص	۱۳۸۷	۱۳۸۸	۱۳۸۹	۱۳۹۰	۱۳۹۱
اعتبار آزمون	۰/۵۲-۰/۸۴	۰/۹۳-۰/۹۵	۰/۹۲-۰/۹۴	۰/۹۰-۰/۹۴	۰/۹۰-۰/۹۴
ضریب دشواری	۰/۵۴-۰/۶۱	۰/۵۸-۰/۶۵	۰/۵۹-۰/۶۵	۰/۶۱-۰/۶۹	۰/۶۱-۰/۶۹
ضریب تمیز	۰/۱۲-۰/۱۸	۰/۲۹-۰/۳۵	۰/۲۷-۰/۳۲	۰/۲۴-۰/۳۱	۰/۲۴-۰/۳۰

جدول ۲: فاصله اطمینان ۹۵٪ برای شاخص‌های کیفی آزمون‌های ارتقاء دستیاری دانشگاه علوم پزشکی کرمانشاه به تفکیک سال

دانشکده / شاخص	**۱۳۸۷**	**۱۳۸۸**	**۱۳۸۹**	**۱۳۹۰**	**۱۳۹۱**
تاکسونومی I	۲۳/۹۷-۴۲/۶۳	۱۷/۶۲-۳۰/۰۹	۲۸/۱۱-۴۲/۸۴	۲۶/۱۷-۳۹/۲۱	۳۵/۱۱-۴۸/۶۵
تاکسونومی II	۳۴/۶۱-۴۴/۷۷	۲۷/۱۳-۳۹/۲۷	۳۴/۳۶-۴۴/۶۳	۳۰/۵۶-۴۳/۹۰	۳۱/۴۵-۳۹/۷۲
تاکسونومی III	۲۱/۶۵-۳۲/۳۴	۳۳/۰۹-۵۲/۷۶	۱۷/۶۴-۳۲/۳۹	۱۹/۷۹-۴۰/۳۵	۱۵/۱۹-۲۹/۸۵
درصد سؤالات بدون اشکال ساختاری	۱۵/۵۶-۳۴/۷۱	۲۹/۷۹-۴۲/۹۳	۷۱/۱۷-۹۲/۹۷	۷۰/۷۳-۹۲/۰۳	۷۷/۴۱-۹۰/۱۲

بحث

یافته‌های این مطالعه نشان داد که آزمون‌های ارتقاء دستیاری دانشگاه علوم پزشکی کرمانشاه از لحاظ شاخص‌های کمی ضریب اعتبار (KR20)، ضریب دشواری، ضریب تمیز و شاخص‌های کیفی، درصد سؤالات بدون اشکال ساختاری و درصد تاکسونومی II و III سؤالات در محدوده قابل قبولی قرار داشتند. محققی و همکاران شاخص‌های کیفی ۲۵ آزمون رشته دانشنامه تخصصی سال‌های ۸۸-۸۶ را از نظر تاکسونومی III و II سؤالات در سال ۸۸ (۱۵/۵۶±۵۶)، در سال ۸۷ (۱۹/۳۰±۴۵) و در سال ۸۶ (۱۸/۹۱±۳۸/۷) درصد گزارش کردند (۲۲). در مطالعه انجام شده در دانشگاه علوم پزشکی کرمانشاه میانگین درصد تاکسونومی II و III (۱۲/۷۴±۶۶/۵۵) بود. طراحی سؤالات در سطوح تاکسونومی II و III نشان‌دهنده تمرکز بیشتر طراح سؤالات بر یادگیری عمیق‌تر در حیطه شناختی یادگیری است اما مشکل اکثر طراحان سؤال چندگزینه‌ای تمرکز بر سطوح یادگیری اولیه حیطه شناختی است (۱۱). گزارش شده است که در دانشگاه علوم پزشکی مازندران ۷۷/۴ درصد سؤالات در تاکسونومی I و مابقی در تاکسونومی II و III قرار داشتند (۱۷) و درصد سؤالات تاکسونومی II و III آزمون‌های پیشرفت تحصیلی دانشگاه علوم پزشکی کرمانشاه نیز ۶۶/۳۳ ٪ بود (۲۷).

بررسی روند طراحی تاکسونومی سؤالات در طول یک دوره پنج ساله نشان داد که از سال ۱۳۸۷ تا سال ۱۳۹۱ میانگین سؤالات دارای تاکسونومی I افزایش و از میانگین سؤالات تاکسونومی II و III کم شده است که این روند از لحاظ آماری معنی‌دار بود(P=۰/۰۰۱).

عدم توجه کافی به رعایت استانداردهای تاکسونومی سؤالات آزمون‌های ارتقاء دستیاری، که بر یادگیری در طبقه-بندی اهداف آموزشی (طبقه بندی بلوم) دلالت می‌نماید، در مطالعات برخی از دانشگاه‌های علوم پزشکی کشور تایید شده است، مانند مطالعه ای که در دانشگاه علوم پزشکی اهواز درباره کمبود توجه به تاکسونومی سطوح دوم و سوم در سؤالات دستیاری انجام شد (۱۷). البته گزارش‌هایی نیز از رعایت استانداردهای تاکسونومیک در حد استاندارد در طراحی آزمون‌های ارتقاء دستیاری نیز وجود دارد (۲۹). کمبود استفاده از سؤالات در سطوح تاکسونومی II و III نه تنها در آزمون‌های ارتقاء دستیاری بلکه در آزمون‌های رشته پرستاری

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

(۲۸) و در سؤالات هیجده درس تخصصی نظری مامایی طی دو نیمسال تحصیلی ۸۶-۱۳۸۵ نیز گزارش شده است (۳۰)، لذا اطلاع‌رسانی به مدیران گروه‌های آموزشی و اعضاء هیئت علمی و در صورت نیاز آموزش آنها، برای اختصاص بیشتر سؤالات آزمون‌ها به سطوح تاکسونومی II و III ضروری به نظر می‌رسد. این راهکار در مطالعه خوشرنگ و همکاران (۱۳۹۲) در مورد بهبود کیفیت سؤالات آزمون‌های ارتقاء دستیاری بعد از مداخله آموزشی در اساتید هیئت علمی دانشگاه گیلان گزارش شده است (۳۱).

یافته‌های این مطالعه نشان داد که میانگین فراوانی سؤالات بدون اشکال ساختاری (۲۹/۴۸٪±۶۲/۵۹٪ است که مشابه نتایج مطالعه محققی و همکاران در بررسی سؤالات آزمون‌های ارتقاء دستیاری پزشکی بود (۲۲). درصد سؤالات بدون اشکال ساختاری در دانشگاه علوم پزشکی مازندران ۴۶٪ بود و پیشنهاد شده است که با تدابیر اصلاحی توزیع تاکسونومی سؤالات و رعایت اصول ساختاری طراحی سؤالات ارتقاء یابند (۱۷).

بررسی روند طراحی سؤالات بدون اشکال ساختاری در طول دوره پنج ساله از سال ۱۳۸۷ تا سال ۱۳۹۱ نشان داد که در ۳۷ آزمون (۵۸/۷٪) بیشتر از ۸۰٪ سؤالات بدون اشکال ساختاری است و تفاوت آماری معنی‌داری در روند پنج ساله این آزمون‌ها دیده نشد. این یافته ضرورت رعایت اصول طراحی سؤالات چندگزینه‌ای میلمن در تهیه سؤالات آزمون‌های ارتقاء توسط اعضاء هیئت علمی را مورد تأکید قرار می‌دهد.

در کل آزمون‌های بررسی‌شده، میانگین ضریب دشواری (۰/۶۲±۰/۰۶۴)، میانگین ضریب تمیز (۰/۲۷±۰/۰۷۳)و میانگین ضریب اعتبار (۰/۸۸±۰/۱۳) بود. بررسی روند ضریب دشواری در طول دوره پنج ساله نشان داد که تفاوت آماری معنی‌داری در شاخص ضریب دشواری وجود دارد به این صورت که در طول سال ۱۳۸۷ الی ۱۳۹۱ ضریب دشواری بالاتر رفته است و آزمون‌ها از لحاظ ضریب دشواری به سوی آسان‌تر شدن هدایت شده‌اند. این روند در شاخص کمی اعتبار کل آزمون (KR20) و در شاخص ضریب تمیز مشاهده شد.

آزمون‌های مطالعه حاضر در محدوده قابل قبول‌تری نسبت آزمون‌های دانشکده پزشکی کوالالامپور (۲۰) از نظر ضریب دشواری بودند. در زمینه اعتبار آزمون (ضریب پایایی آزمون) آنالیز ۲۴ آزمون دانشکده پزشکی مشهد نشان داد که میانگین اعتبار آزمون‌ها ۰/۷۲ و در ۵۲/۲ درصد سؤالات ضریب دشواری نامناسب و در ۴۹/۲ درصد موارد ضریب تمیز قابل‌قبول نبودند (۱۹). در دانشکده پزشکی قزوین نیز تنها ۱۶/۳ درصد سؤالات از نظر ضریب دشواری و تمیز مناسب گزارش شده است (۱۵). در خارج از ایران نیز یافته‌های مطالعه باچوف و همکاران در بررسی سؤالات دانشگاه مکزیکوسیتی میانگین ضریب دشواری سؤالات را ۰/۳۴ و میانگین ضریب تمیز را ۰/۲۸ اعلام کرده‌اند (۳۲). پایین بودن ضریب تمیز کمتر از محدوده استاندارد در آزمون‌های رشته پرستاری (۲۸) و مامایی (۳۰) نیز گزارش شده است.

از محدودیت‌های این مطالعه می‌توان به عدم مقایسه نتایج در بین اساتید درون هر گروه و تفاوت‌های دانشجویان پزشکی در سال‌های مختلف اشاره کرد.

نتیجه‌گیری

با توجه به این نتایج و توجه به ارتقاء شاخص‌های کمی در طراحی آزمون‌های ارتقاء دستیاری و کاهش کیفیت سؤالات در شاخص های تاکسونومی II و III پیشنهاد می‌شود با آموزش و تقویت بنیه علمی اساتید و مدرسین طراح آزمون‌های ارتقاء دستیاری نسبت به ارتقاء دانش آنها در مفاهیم کیفی سؤالات و حیطه‌های یادگیری اقدام شود.

در نهایت پیشنهاد می‌شود در آینده مطالعاتی در مورد بررسی ارتباط شاخص‌های کمی و کیفی سؤالات برای ارزیابی، تأثیر افزایش شاخص‌های کمی و تأثیر بر کیفیت سؤالات، بررسی ارتباط خصوصیات فردی دستیاران مانند ضریب هوشی /هوش هیجانی، میزان علاقه‌مندی به رشته و جنس بر شاخص‌های کمی و کیفی آزمون‌های ارتقاء در جهت روش شدن ابعاد این موضوع انجام گردد.

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

سپاسگزاری

این مقاله گزارش نهایی طرح تحقیقاتی با همین عنوان و با حمایت‌های مالی معاونت تحقیقات و فن آوری دانشگاه علوم پزشکی کرمانشاه به تاریخ ۱۳۹۱/۷/۱۲ و شماره ۹۱۲۰۱ است. بدین وسیله از ریاست و واحدهای ارزشیابی و پژوهش در آموزش مرکز مطالعات و توسعه آموزش پزشکی معاونت آموزشی دانشگاه علوم پزشکی کرمانشاه جهت راهنمایی‌های ارزشمند، نهایت تشکر به عمل می‌آید.

References

1– ShabaniH. Instructional skills, methods and techniques of teaching.15thed. Ghom:Samt. 2007: 367. [Persian]

2– Saif AA. Educational measurement, assessment and evaluation. 2th ed.Tehran: Doran.2003:27,28,32. [Persian]

3– Mohmoodi SM.(In translation)A Handbook for medical teachers:Newble D, Cannon R. 1st ed. Tehran Health Ministry Publisher. 1997:129-30. [Persian]

4– Asetzadeh S.The good assessment guide. 1sted. Ghazvin:Bahrololum. 2001:41,48. [Persian]

5– Zolfagari B,Adibi N, Drakhshonfor S, et al. Academic achvementtests. 1sted.Isfahan:Neshat. 2001: 20,41,43,61 . [Persian]

6– Rezaii H, ShogiF, Dolati Y, et al. Measurement and evaluation in medicine. 1sted. Tehran:Ameh. 1998:75. [Persian]

7- Epstein RM.Assessment in medical education. The New England Journal of Medicine. 2007;356(4):387-96.

8- Ben-David MF.AMEE Guide No. 18: Standard setting in student assessment. Medical Teacher.2000; 22(2): 120-30.

9- Larsen DP, Butler AC, Roediger HL.Test-enhanced learning in medical education.Medical Education.2008;42(10): 959-66.

10- Reid WA, Duvall E, Evans P.Relationship between assessment results and approaches to learning and studying in Year Two medical students. Medical Education. 2007; 41(8): 754-62.

11- Brady AM.Assessment of learning with multiple-choice questions.Nurse Education inPractice.2005; 5(4): 238-42.

12- Roshanpour F, PourmirzaKalhori R, Rezai M. Quantitiveanalysis of medical promotion exams in Kerman shah Health and Education University in 2008. Special supplement for 10th National Congress of Medical Education – May.2009; 335-6. [Persian]

13 – Grakyaraghi M, Toghyanifar N, Esmaili A, et al. Evaluation of MGQ on clinical trianing in Esfahan Medical University 2005 – 2006. Strides in Develoment of Medical Education. 8th National Congress of Education – March. 2007; 73. [Persian]

14 –Javadi M, Abbaszade A, Borhani F, et al. Analytic evalution of MCQ in end termisterexams in nursing school of Baam.Strides in Development

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

of MedicalEducation .8th National Congress of Medical Eduction – March 2007; 217. [Persian]

15– Drakhsan F, Jahanihashemi H. Evaluation of end term exams of GhazvinMedical University in 2007-2008. Special supplement.10th National Congress of Medical Education – May 2009: 308-9. [Persian]

16- Kaveh Tabatabaee MS, Bahreyni Toosi MH, Derakhshan A,et al.Analytic assessment of multiple-choice tests.Journal of Medical Education. 2003; 2(2):87-91.

17– Latifi SM, Shakornia A. Comparision of experts behaviors and WWievs about MCQ designedin Ahvaz Medical University. Strides in Development of Medical Education. 8th National Congress of Medical Education – March 2007: 159-60. [Persian]

18– Haghshenas MR, Vahidshahi K.Mahmudi M, et al.Evaluation of multiple choice questions in the School of Medicine, Mazandaran University of Medical Sciences,the first semester of 2007.Strides in Development of Medical Eduction. 2008; 5(2):119-26 . [Persian]

19– Pourmirza Kalhori R, Roushanpour F, Rezaei M,Esmaeili K. Analysis of multiple choice tests designed byfaculty members of Kermanshah University of MedicalSciences. Educ Res Med Sci. 2013; 2(2):18-19.

20-Mitra NK, Nagaraja HS, Ponnudurai G, et al. The levels of difficulty and discrimination indices in type Amultiple choice questions of pre-clinical semester 1multidisciplinary summative tests.International E-Journal of Science, Medicine and Education. 2009; 3(1): 2-7.

21-Quadrelli S, Davoudi M, Galíndez F, et al. Reliability of a 25-item low-stakes multiple-choice assessment of bronchoscopic knowledge.Chest. 2009; 135(2):315-21.

22- Mohagheghi MA, Vahidshahi K, Shakeri S, et al. Evaluation of some quality index of multiple choice questions in the board exams in medicine from 2005- 2007. [Cited.Sepmter,93.2011]. available from:Http://cgme.behdasht.gov.ir.

23- PourmirzaKalhori R, Razazian N, Naderipour A, et al.Educational purpose and medical students' achievements in clinical training in an university hospital in Kermanshah.Journal of Kermanshah University of Medical Sciences (Behbood).2010; 13(4): 344-52. [Persian]

24- PaulinoAC, Kurtz E. American College of Radiology in-training examination for residents in radiation oncology.International Journal of Radiation Oncology*Biology*Physics. 2008; 70(3): 666-70.

25- Hammond EJ, McIndoe AK, Sansome AJ, et al. Multiple-choice examinations: Adopting an evidence-based approach to exam technique. Anaesthesia. 1998; 53(11): 1105-8.

26- McCoubrie P. Improving the fairness of multiple-choice questions: A literature review. Medical Teacher.2004;26(8): 709-12.

27- Pour mirzakalhoriR. Analytic assessment of multiple choice questions (MCQ) of faculty members' scholarship in Kermanshah Medical Sciences University in 2010.Final report of

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

research proposal accepted in deputy of research in Kermanshah Medical University in 90016 number and date 1390/2/20. [Persian]

28– ShabanM, RamezaniBadr F.Effect of test item analysis of summative exams on quality of test designing.Hayat. 2007; 13(1): 5-15. [Persian]

29– Roshanpour F,PourmirzaKalhori R, Rezai M. Quantitiveanalysis of medical promotion exams in Kermanshah health and education University in 2008. Special Supplement for 10th National Congress of Medical Education– May 2009: 335-36. [Persian]

30- Kazemi A,Ehsanpour S,Hassanzadeh A. Investigating the academic achievement evaluation of specialized theoretical courses of midwifery BS.Iranian Journal of Medical Education. 2010;9(4): 346-55 . [Persian]

31- Taheri M, Khoshrang H, Asadi A, et al. Quality of residents' promotion exams before and after educational intervention in 2010-2011.Iranian Journal of Medical Education. 2013; 13(7): 551-60. [Persian]

32- Backhoff E, Larrazolo N, Rosas M.The level of difficulty and discrimination power of the basic knowledge and skills examination (EXHCOBA).Electronic Journal of Educational Research.2000; 2(1):1-16.

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]

A survey of quality and quantity indexes of multiple choice question (MCQ) exams of medical residents at Kermanshah University of Medical Sciences: 2008-2012

Pourmirza kalhori R (MSc.)[1], Rezaei M (PhD)[2], Karami Matin B (PhD)[3], Roshanpour F (MSc) *[4]

1- Dept. of Emergency Medicine, School of Paramedics, Kermanshah University of Medical Sciences, Kermanshah, Iran
2- Dept. of Biostatistics, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran
3- Dept. of Republic Health, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran
4- Deputy of research and technology in Kermanshah University Medical Sciences, Kermanshah University of Medical Sciences, Kermanshah, Iran

Received: 18 Agu 2013 ***Accepted: 9 Dec 2013***

Abstract

Introduction: Multiple choice questions (MCQs) are widely used to assess medical residents. The study aims to analyse MCQ exams of medical residents administered at Kermanshah University of Medical Sciences during 2008-2012.

Method: Data of this retrospective study were extracted from 63 MCQ exams including 9600 tests. Quantity related variables included; discrimination index, difficulty level, reliability of test index (KR20). Quality related variables included; proportion of questions without structural defect and proportion of taxonomy 1, 2 and 3 questions. Data were collected using MCQs from question bank and analyzed by SPSS V.16 software.

Results: Over the specified five years, the average difficulty level was %0.62, the average discrimination index was 0.27 and the average of reliability coefficient (KR20) in total exams was 0.88. The average of questions in Taxonomy 1 over the specified years was 33.5% and questions with taxonomy 2 and 3 were66.5%. The average of questions without structural defect was 62.6% which fell in the unacceptable range. A significant difference was observed in reliability coefficient (KR20)- (P=0.017), difficulty level (P=0.001), taxonomy 1 (P=0.001) and Taxonomy 2 and 3 (P=0.006).

Conclusion: According to findings of this study, taxonomy 2 and 3 were decreased and taxonomy 1 was increased over the specified five years. We recommend the specialist board members and the Deputy for Education of the Ministry of Health to consider the results of this study and react to them by improving MCQ quality of future exams.

Keywords: Multiple choice questions, exam, analysis, l residency, board

*Corresponding author's email: roshanpour_f@yahoo.com

This paper should be cited as:
Pourmirza kalhori R, Rezaei M, Karami Matin B, Roshanpour F. ***A survey of quality and quantity indexes of multiple choice question (MCQ) exams of medical residents at Kermanshah University of Medical Sciences: 2008-2012.*** Journal of Medical Education and Development. 2014; 8(4): 64-75

[Downloaded from jmed.ssu.ac.ir on 2026-11-19]