

## تحلیل کمی و کیفی آزمون های چند گزینه ای اعضای هیات علمی دانشگاه علوم پزشکی کرمانشاه در سال ۱۳۹۰

### چکیده

**زمینه:** آزمون های چند گزینه ای یکی از روش های ارزشیابی در علوم پزشکی است. این مطالعه با هدف تجزیه و تحلیل کمی و کیفی آزمون های چند گزینه ای اعضای هیات علمی دانشگاه علوم پزشکی کرمانشاه از نظر کمی و کیفی در سال ۱۳۹۰ انجام شد.

**روش ها:** در این مطالعه توصیفی-تحلیلی ۱۵۶ آزمون از لحاظ شاخص های کمی ضریب دشواری، ضریب تمیز و اعتبار کل آزمون سوالات و شاخص های کیفی درصد تاکسونومیک I، II، III و درصد سوالات بدون اشکال ساختاری آنالیز شدند. داده ها با نرم افزار SPSS و آزمون های آزمون t و مجذور کای و ANOVA و آزمون آماری مقایسات چند گانه توکی (Tukey) تجزیه و تحلیل شد.

**یافته ها:** میانگین ضریب اعتبار کل آزمون ها  $0/31 (\pm 0/68)$ ، میانگین ضریب دشواری  $0/15 (\pm 0/56)$ ، میانگین ضریب تمیز  $0/21 (\pm 0/15)$ ، میانگین انحراف معیار نمرات  $4/38 (\pm 5/96)$  بود. در شاخص های کیفی میانگین درصد سوالات بدون اشکال ساختاری کل آزمون ها  $14/88 (\pm 73/88)$ ، میانگین تاکسونومی I  $15/78 (\pm 34/65)$ ، میانگین تاکسونومی II  $13/14 (\pm 41/34)$  و میانگین تاکسونومی III  $15/40 (\pm 23/99)$  بود. آزمون آماری مقایسات چند گانه توکی (Tukey) تفاوت آماری معنی دار بین اعتبار کل آزمون های دانشکده پزشکی با سایر دانشکده ها نشان داد ( $P=0/001$ ).

**بحث و نتیجه گیری:** نتایج این مطالعه نشان داد که ضریب تمیز، درصد تاکسونومی II و III سوالات و درصد سوالات بدون اشکال ساختاری آزمون های طراحی شده اعضا هیات علمی دانشگاه علوم پزشکی کرمانشاه کمتر از مبنای قابل قبول است، لذا پیشنهاد می شود با تقویت بنیه علمی اساتید نسبت به ارتقا دانش آنها در طراحی آزمون های چند گزینه ای اقدام نمود.

**کلید واژه ها:** سوالات چند گزینه ای، اعتبار آزمون، ضریب دشواری، ضریب تمیز

### رضا پور میرزا کلهری<sup>۱</sup>،

فاطمه دارابی<sup>۲</sup>، فرح روشن پور<sup>۳\*</sup>،  
منصور رضایی<sup>۴</sup>

۱. گروه فوریت های پزشکی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.

۲. گروه هوشبری، دانشکده پیراپزشکی، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.

۳. معاونت تحقیقات و فناوری، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

۴. گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.

\* **عهده دار مکاتبات:** واحد ارزشیابی، مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.

E-mail: roshanpour\_f@yahoo.com

### مقدمه

در فرآیند آموزش، هدف ایجاد یادگیری در فراگیران است<sup>۱</sup>. و ابزار ارزیابی وقوع یادگیری در فراگیران آزمون است<sup>۲</sup>. به دلیل نقش مهم آزمون در سنجش یادگیری،

ضروری است که هر آزمون دارای سه شاخص روایی، پایایی و عملی بودن باشد<sup>۳</sup>. آزمون ها در دو دسته کلی آزمون های معلم ساخته و آزمون های استاندارد طبقه بندی می شوند<sup>۴</sup>. دسته بندی دیگر آزمون ها از

بودن اعتبار آزمون ها<sup>۸</sup>، کم بودن درصد سوالات قابل قبول از نظر ضریب دشواری و تمیز<sup>۹</sup> وجود دارد. نتایج این مطالعات به قدری مورد توجه است که اساتید و کارشناسان خبره طراحی سوالات چند گزینه ای اذعان می دارند که اشکالاتی در زمینه طراحی این گونه سوالات موجود است<sup>۱۰</sup>. این ایرادات کمی و کیفی فقط اختصاص به آزمون های اختصاصی رشته پزشکی ندارند. برای نمونه ارزشیابی آزمون های کتبی چند گزینه ای درس بهداشت دانشجویان پرستاری و سوالات مقطع استازری پزشکی منعکس کننده اشکالات مهم ساختاری در طراحی سوالات است<sup>۱۱</sup>. با این رویکرد بود که آنالیز سوالات مدرسین علوم پزشکی در دستور کار مراکز توسعه و آموزش علوم پزشکی کشور قرار گرفت و این بررسی یکی از شاخص های ارزیابی عملکرد سالیانه مراکز توسعه و آموزش علوم پزشکی کشور از لحاظ رتبه بندی کشوری دانشگاه های علوم پزشکی کشور می باشد. مطالعات مختلفی پیشنهاد کننده آموزش اساتید علوم پزشکی در زمینه طراحی آزمون های چند گزینه ای هستند<sup>۱۲</sup>. البته نباید از ذهن دور داشت که این مشکلات اختصاص به کشور ایران ندارد بلکه در سایر کشور ها نیز گزارشات متعددی مبنی بر نامناسب بودن شاخص های کمی و کیفی آزمون ها وجود دارد<sup>۱۳،۱۴</sup>. با این پیش فرض ها، هدف کلی این مطالعه تحلیل آزمون های چند گزینه ای طراحی شده توسط اعضا هیات علمی دانشگاه علوم پزشکی کرمانشاه از نظر کمی و کیفی در سال ۱۳۹۰ بود.

لحاظ عینیت در تصحیح آنها است که این خصوصیت آزمون را به دو نوع عینی و شفاهی تقسیم می نماید. منظور از آزمون عینی آزمونی است که در صورت تصحیح مکرر و توسط افراد مختلف دارای نتیجه ای یکسان باشد. از آزمون های عینی می توان آزمون های صحیح- غلط و چند گزینه ای را نام برد<sup>۵</sup>. شایع ترین سوالات عینی مورد استفاده سوالات چند گزینه ای هستند. هر سوال دارای بخش ریشه و گزینه های پیشنهادی است که یکی از آنها پاسخ صحیح است و فراگیر باید پاسخ صحیح را از بین گزینه های پیشنهادی/ انحرافی انتخاب کند<sup>۴</sup>. سوالات چند گزینه ای در حیطه شناختی دانش، فهم، ادراک، کاربرد و اطلاعات را ارزشیابی می کنند<sup>۶</sup>. به طور معمول ضروری است که بعد از اجرای هر آزمون تجزیه و تحلیل سوالات از لحاظ کمی و کیفی انجام شود. بُعد کیفی آنالیز سوالات منطبق بر اهداف آموزشی بنیامین بلوم و شامل: تاکسونومی I (سطوح دانشی)، تاکسونومی II (یاد آوری) و تاکسونومی III (کاربرد مفاهیم) و تعیین درصد سوالات بدون اشکال ساختاری است. در بُعد کمی اعتبار کل آزمون و ضریب های دشواری و تمیز آزمون سنجش می شود. در سالیان اخیر این کار با استفاده از نرم افزارهایی که با استفاده از نظریه کلاسیک آزمون ها Classical Test Theory (CTT) برنامه ریزی شده اند تحلیل کمی آزمون ها انجام می شود. مطالعات مختلفی در ایران در زمینه آموزش علوم پزشکی به تحلیل آزمون های چند گزینه ای معلم ساخته پرداخته اند که نتایج آنها قابل توجه است. گزارشاتی از پایین بودن ضریب های دشواری (سخت بودن آزمون)<sup>۷</sup>، آسان بودن سوالات و مناسب

## مواد و روش ها

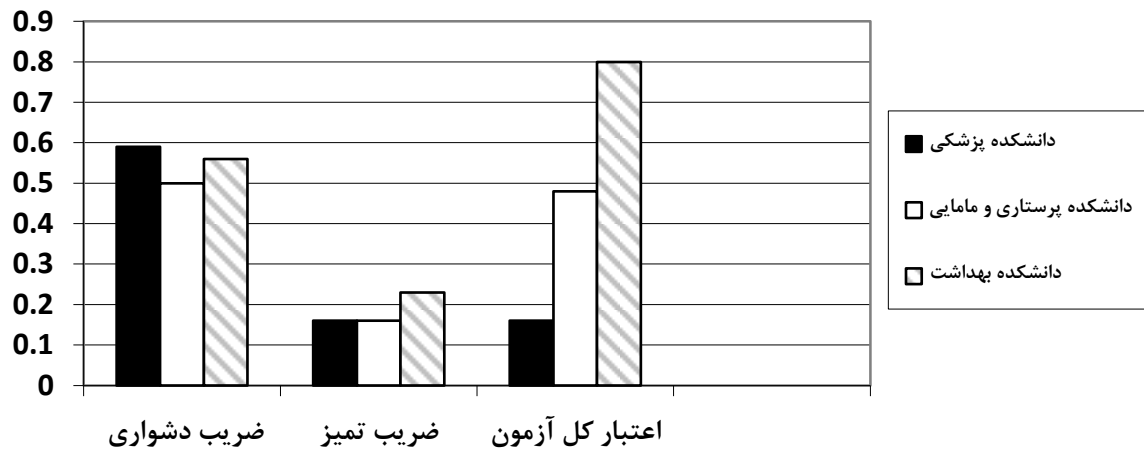
تمییز بالاتر از ۰/۳ مناسب، صفر تا ۰/۳ کم و زیر صفر ضعیف طبقه بندی شدند<sup>۱۶</sup>. باید در نظر داشت نزدیک شدن ضریب دشواری به عدد یک نشانه آسان بودن آزمون و نزدیک شدن به صفر نشانه سخت شدن آزمون است و در مورد ضریب تمییز نزدیک شدن ضریب به عدد یک نشانه معتبر بودن آزمون در تفکیک دانشجویان زرنگ از دانشجویانی است که پاسخ صحیح به سوالات نداده اند. رعایت اختصاص حدود ۲۰٪ سوالات در تاکسونومی I و ۸۰٪ سوالات به تاکسونومی های II و III در این مطالعه به عنوان رعایت استاندارد در شاخص های حیطه شناختی سوالات و رعایت ۸۰ درصد سوالات بدون اشکال ساختاری مبنای درصد مناسب سوالات بدون اشکال ساختاری بود. جهت رعایت ملاحظات اخلاقی و حفظ شئون معلمی به اسامی و گروه های آموزشی مدرسین اشاره نشد و مقایسه مدرسین نیز با یکدیگر انجام نشد. برای انجام این پژوهش مجوز های لازم از ریاست مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی کرمانشاه اخذ شد.

## یافته ها

در این مطالعه از مجموع ۱۵۶ آزمون معلم ساخته مورد بررسی، ۸۲ آزمون (۵۲/۶٪) از دانشکده پزشکی، ۳۳ آزمون (۲۱/۱٪) از دانشکده پیراپزشکی، ۲۴ آزمون (۱۵/۴٪) از دانشکده پرستاری و مامایی، ۱۵ آزمون (۹/۶٪) از دانشکده بهداشت و ۲ آزمون (۱/۳٪) از دانشکده داروسازی طراحی شده بودند. میانگین ضریب اعتبار کودریچاردسون (KR ۲۰) آزمون ها (۰/۶۸ ± ۰/۳۱)، میانگین ضریب دشواری کل آزمون ها (۰/۵۶ ± ۰/۱۵) و میانگین ضریب تمییز کل آزمون ها (۰/۲۱ ± ۰/۱۵) بود، نمودار (۱).

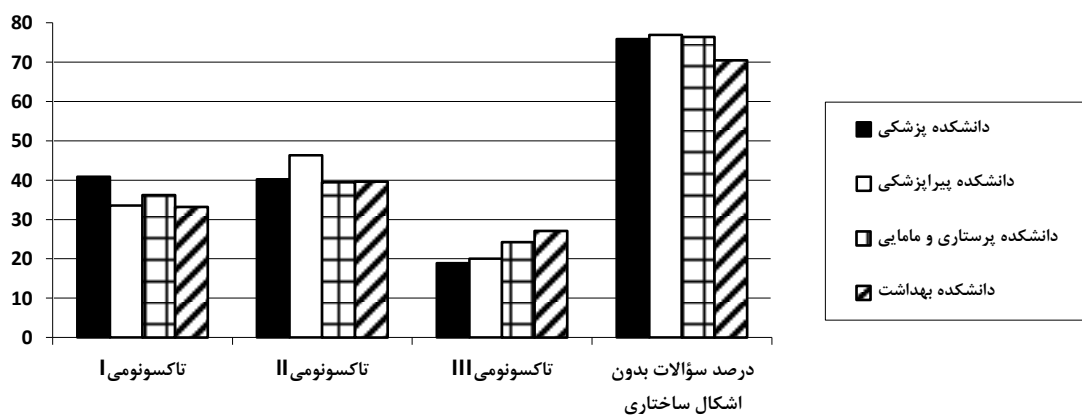
این مطالعه توصیفی- تحلیلی با استفاده از گزارشات تحلیلی آزمون های اعضا هیات علمی دانشگاه علوم پزشکی کرمانشاه که در سال تحصیلی ۹۰-۱۳۸۹ نمونه سوالات چند گزینه ای خود را برای تحلیل کمی و کیفی به واحد ارزشیابی مرکز مطالعات و توسعه آموزش علوم پزشکی دانشگاه علوم پزشکی کرمانشاه ارسال کرده بودند، انجام شد. روش نمونه گیری به صورت تمام شماری بود و بر اساس اطلاعات موجود ۱۵۶ آزمون از این مدرسین مورد بررسی قرار گرفت و حجم نهایی نمونه را تشکیل دادند.

ابزار جمع آوری داده ها گزارشات کمی و کیفی تحلیل آزمون های چند گزینه ای بود. برای آنالیز کمی آزمون ها از نرم افزار یگانه تحت برنامه ویندوز XP استفاده شد و شاخص های کمی ضریب دشواری، ضریب تمییز و اعتبار کل آزمون از گزارشات آن استخراج شد. برای جمع آوری داده های کیفی تحلیل آزمون ها از گزارشات مرکز مطالعات و توسعه آموزش پزشکی شاخص های درصد تاکسونومی I و II و III سوالات و درصد سوالات بدون اشکال ساختاری استفاده شد. برای تجزیه و تحلیل داده ها از نرم افزار SPSS نسخه ۱۸ استفاده و برای خلاصه سازی داده ها از محاسبه میانگین و انحراف معیار شاخص ها و تعیین درصد ها، جداول دو بعدی به صورت تعداد و درصد و آزمون t، مجذور کای، ANOVA و آزمون آماری مقایسات چند گانه توکی (Tukey) استفاده شد. بر اساس مطالعه وفا مهر و دادگستر نیا (۱۳۸۹) ضریب دشواری ۰/۷-۰/۳ مناسب، بیشتر از ۰/۷ آسان و کمتر از ۰/۳ سخت و ضریب



شکل ۱: میانگین شاخص های کمی آزمون های دانشکده های اقماری دانشگاه علوم پزشکی کرمانشاه

در بررسی شاخص های کمی کل آزمون ها میانگین درصد سوالات بدون اشکال ساختاری (  $0.73/0.88 \pm 0.14/0.88$  )، میانگین تاکسونومی I سوالات کل آزمون ها (  $0.34/0.65 \pm 0.15/0.78$  )، میانگین تاکسونومی II سوالات کل آزمون ها (  $0.41/0.34 \pm 0.13/0.14$  ) و میانگین تاکسونومی III سوالات کل آزمون ها (  $0.23/0.99 \pm 0.15/0.40$  ) بود، نمودار (۲).



شکل ۲: میانگین شاخص های کیفی آزمون های دانشکده های اقماری دانشگاه علوم پزشکی

آزمون ها در دانشکده پرستاری و مامایی بود ( $P < 0.001$ ) و میانگین اعتبار آزمون ها در دانشکده پرستاری و مامایی کمتر از میانگین اعتبار آزمون ها در دانشکده پرستاری و مامایی بود ( $P = 0.005$ ).

براساس آزمون آماری مقایسات چند گانه توکی (Tukey) میانگین اعتبار آزمون ها در دانشکده پرستاری و مامایی کمتر از میانگین اعتبار آزمون ها در دانشکده پزشکی بود ( $P = 0.019$ ). میانگین اعتبار آزمون ها در دانشکده بهداشت کمتر از میانگین اعتبار

بهداشت بود. آزمون آماری مقایسات چند گانه توکی (Tukey) نشان داد که میانگین ضریب تمییز دانشکده پرستاری و مامایی کمتر از میانگین ضریب تمییز آزمون های دانشکده پزشکی بود ( $P=0/002$ ). میانگین ضریب تمییز آزمون های دانشکده بهداشت کمتر از میانگین ضریب تمییز آزمون های دانشکده پیراپزشکی بود ( $P=0/024$ ). با توجه به اهمیت مشخص شدن فاصله اطمینان شاخص های کمی و کیفی آزمون های دانشکده های مورد مطالعه و مقایسه این شاخص ها با موارد توصیه شده توسط منابع علمی آموزش پزشکی این فواصل اطمینان در جداول ۱ و ۲ ذکر شده اند.

بالا ترین ضریب دشواری (سخت ترین) آزمون برابر  $0/13$  در دانشکده پیراپزشکی و پایین ترین برابر  $0/87$  (آسانترین) نیز در این دانشکده بود. آزمون آماری مقایسات چند گانه توکی (Tukey) نشان داد که میانگین ضریب دشواری دانشکده پیراپزشکی کمتر از میانگین ضریب دشواری آزمون های دانشکده بهداشت بود ( $P=0/007$ ). میانگین ضریب دشواری آزمون های دانشکده پزشکی کمتر از میانگین ضریب دشواری آزمون های دانشکده پرستاری و مامایی بود ( $P<0/001$ ). بالاترین ضریب تمییز آزمون، برابر  $0/9$  در دانشکده های پزشکی و پیراپزشکی و پایین ترین برابر  $0/09$  در دانشکده

جدول ۱: فاصله اطمینان شاخص های کمی آزمون های دانشکده های دانشگاه علوم پزشکی کرمانشاه

شاخص	دانشکده	بهداشت	پیراپزشکی	پرستاری و مامایی	پزشکی
اعتبار آزمون		$0/33-0/6$	$0/52-0/69$	$0/24-0/73$	$0/75-0/85$
ضریب دشواری		$0/48-0/68$	$0/54-0/68$	$0/41-0/52$	$0/53-0/58$
ضریب تمییز		$0/08-0/23$	$0/12-0/28$	$0/08-0/28$	$0/2-0/26$

ضریب تمییز آزمون های دانشکده بهداشت کمتر از میانگین ضریب تمییز آزمون های دانشکده پیراپزشکی بود ( $P=0/014$ ). بیشترین درصد سوالات بدون اشکال ساختاری برابر  $98/5\%$  در دانشکده های پزشکی و کم ترین برابر  $45\%$  در دانشکده بهداشت بود. جدول (۲) فواصل اطمینان شاخص های کیفی بررسی شده آزمون ها را نشان می دهد.

بیشترین درصد تاکسونومی I و II و III در سوالات آزمون های دانشکده های پزشکی، کمترین تاکسونومی I در دانشکده پیراپزشکی و کمترین تاکسونومی II در دانشکده بهداشت وجود داشت. آزمون آماری مقایسات چند گانه توکی (Tukey) نشان داد که میانگین تاکسونومی III دانشکده پرستاری و مامایی کمتر از میانگین ضریب تمییز آزمون های دانشکده پزشکی بود ( $P=0/012$ ). میانگین

جدول ۲: فاصله اطمینان شاخص های کیفی آزمون های دانشکده های دانشگاه علوم پزشکی کرمانشاه

شاخص	دانشکده	بهداشت	پیراپزشکی	پرستاری و مامایی	پزشکی
تاکسونومی I		$31/16-50/57$	$27/71-39/40$	$29/60-42/73$	$29/74-36/79$
تاکسونومی II		$29/91-50/59$	$42/15-50/56$	$34/82-44/16$	$36/72-42/64$
تاکسونومی III		$11/87-25/86$	$15/82-24/32$	$17/46-31/10$	$23/22-30/86$
درصد سوالات بدون اشکال ساختاری		$14/9-82/59$	$72/49-81/46$	$71/21-81/64$	$66/87-74/13$

دشواری و ضریب تمیز در دانشگاه علوم پزشکی کرمانشاه است<sup>۱۳</sup>. تحلیل سوالات چند گزینه‌ای دروس های آناتومی، فیزیولوژی، بیوشیمی، ژنتیک، آمار و علوم رفتاری در دانشکده پزشکی کوالا لامپور مالزی<sup>۱۴</sup> و تحلیل سوالات دانشکده پزشکی بوینوس آیرس آرژانتین<sup>۱۵</sup> نشان داد که در شاخص ضریب دشواری و ضریب تمیز هر دو دانشکده، در محدوده قابل قبولی هستند. گزارش شده است که بهترین ضریب تمیز سوال در زمانی است که سوال دارای درجه دشواری متوسط باشد و با افزایش درجه دشواری سوال ضریب تمیز سوال منفی و نامناسب می شود<sup>۲۰</sup>. نکته ای که باید به آن توجه داشت این است که در تفسیر داده های حاصل از تجزیه و تحلیل آزمون ها باید احتیاط کرد زیرا ضریب تمیز یا تفاوت عملکرد دو گروه قوی و ضعیف از دانشجویان همیشه بیان کننده کیفیت سوالات نمی باشد و نباید این سوالات از بانک سوالات حذف شوند بلکه باید به دقت بررسی شوند و به نقش آنها در اندازه گیری دستیابی به اهداف آموزشی توجه داشت. در حقیقت بودن سوالاتی که دارای درجه دشواری آسان یا بسیار سخت هستند که دارای درجه تمیز پایینی می باشند و در آزمون ها وجود آنها برای دستیابی به نمونه کاملی از محتوا و اهداف آموزشی ضروری است<sup>۲۱</sup>. با وجود گزارشاتی مبنی عدم خطی بودن ارتباط ضریب دشواری و تمیز در سوالات آموزش پزشکی<sup>۲۲</sup> نباید از نظر دور داشت که شاخص های کمی ضریب دشواری و ضریب تمیز آزمون ها دارای ارتباطی معکوس با یکدیگر هستند به این معنی که با افزایش ضریب دشواری (آسان تر شدن سوال) ضریب تمیز سوال کمتر می شود و همچنین با آسان طراحی کردن سوالات، قدرت سوال در تفکیک دانشجویان زرننگ و درس خوان از دانشجویانی که به طور شانس به سوال پاسخ می دهند، کمتر می شود. لذا ضروری به نظر می رسد که مدرسین علوم پزشکی با تحلیل آزمون های خود ارتباط ضریب های دشواری و تمیز سوالات طراحی شده خود را برای طرح سوالات بهینه سنجش نمایند

آزمون آنالیز واریانس شاخص های کمی و کیفی نشان داد که تفاوت آماری معنی داری بین شاخص های ضریب دشواری، ضریب تمیز، اعتبار کل آزمون و انحراف معیار سوالات و درصد تاکسونومی II آزمون های بررسی شده وجود دارد ( $P < 0/05$ ) ولی در درصد تاکسونومی I و III سوالات تفاوت معنی دار آماری مشاهده نشد.

## بحث

رایج ترین نوع آزمون کتبی در آموزش علوم پزشکی آزمون های چند گزینه‌ای است<sup>۳</sup>. نتایج این مطالعه نشان داد که آزمون های چند گزینه‌ای طراحی شده توسط اساتید هیات علمی دانشگاه علوم پزشکی کرمانشاه از لحاظ شاخص های کمی ضریب اعتبار ( $KR = 20$ )، ضریب دشواری و ضریب تمیز در محدوده قابل پذیرشی بود. اما در شاخص های کیفی مانند: درصد سوالات بدون اشکال ساختاری، تاکسونومی I و II و III سوالات کمتر از سطوح پیشنهاد شده بود.

تحلیل ۲۴ آزمون دانشکده پزشکی مشهد نشان داد که میانگین اعتبار آزمون ها  $0/72$  و در  $52/2\%$  سوالات ضریب دشواری نامناسب و  $49/2\%$  آزمون ها دارای ضریب تمیز قابل قبول نبودند<sup>۱۷</sup>. در دانشکده پزشکی قزوین نیز تنها  $16/3\%$  سوالات از نظر ضریب دشواری و تمیز مناسب گزارش شده است<sup>۹</sup>. نامناسب بودن شاخص کمی ضریب دشواری نه تنها در ایران بلکه در خارج از کشور نیز گزارش شده است که می توان به مطالعه Bachhoff و همکاران (۲۰۰۰) در بررسی سوالات دانشگاه مکزیکوسیتی اشاره کرد<sup>۱۸</sup>.

نتایج مطالعه شبان (۱۳۸۵) در دانشگاه تهران اشاره به پایین بودن ضریب تمیز کمتر از سطح قابل قبول در برخی از آزمون های مدرسین داشت<sup>۱۹</sup>. در شاخص اعتبار آزمون نتایج مطالعه تایید کننده مطالعه انجام شده در دانشگاه علوم پزشکی مازندران بود و تفاوت نتایج آنالیز اساتید دانشگاه علوم پزشکی کرمانشاه با مطالعه انجام شده در دانشگاه علوم پزشکی مازندران، قابل قبول تر بودن شاخص های ضریب

که در سطح تاکسونومی بالاتری قرار دارند را گزارش کرده اند<sup>۱۶</sup>. هر چند که در این مورد Kibble & Johnson (۲۰۱۱) گزارش کرده اند که سطح تاکسونومی سوالات نه تنها با درجه تمییز، بلکه با نمرات دانشجو نیز ارتباط ندارند<sup>۲۷</sup>. به نظر می رسد هنوز به مطالعات بیشتری در علوم پزشکی نیاز است تا ارتباط بین شاخص های کمی و کیفی به درستی شناخته شود.

در این مطالعه شاخص های کمی ضریب تمییز و ضریب دشواری و سطوح تاکسونومی سوالات در دانشکده های مختلف با هم دارای تفاوت آماری معنی داری بودند که مشابه نتایج مطالعه خوشرننگ و همکاران (۱۳۹۲) در بررسی سوالات آزمون های ارتقا دستیاری در گروه های مختلف آموزشی دانشگاه گیلان بود<sup>۲۸</sup>.

### نتیجه گیری

نتایج این مطالعه نشان داد که آزمون های طراحی شده توسط مدرسین دانشگاه علوم پزشکی کرمانشاه، از نظر میانگین ضریب تمییز، درصد تاکسونومی II و III سوالات و درصد سوالات بدون اشکال ساختاری، کمتر از مبنای قابل قبول است. لذا پیشنهاد می شود با تقویت بنیه علمی اساتید نسبت به ارتقاء دانش آنها در طراحی آزمون های چند گزینه ای، به عنوان ابزاری جهت ارزشیابی دانشجویان اقدام کرد.

### تقدیر و تشکر

این مقاله از طرح تحقیقاتی شماره ۹۰۰۱۶ استخراج شده و با حمایت های مالی معاونت تحقیقات و فن آوری دانشگاه علوم پزشکی کرمانشاه انجام شده است.

### References:

1. Shabani H. Instructional skills (methods and techniques of teaching). 8<sup>th</sup> ed. Tehran: SAMT publication; 2003 [Persian].
2. Saif AA. Educational measurements, assessment and evaluation. 7<sup>th</sup> ed. Tehran: Doran Publication; 2010 [Persian].

کمبود توجه به سوالاتی که بر یادگیری در سطوح عالی طبقه بندی اهداف آموزشی (طبقه بندی بلوم) دلالت می نماید در رشته های علوم پزشکی مشهود است. گزارش شده است که در دانشگاه علوم پزشکی مازندران ۷۷/۴٪ سوالات در تاکسونومی I و ۲۲/۶٪ در تاکسونومی II و III طراحی شده است<sup>۱۳</sup> و در دانشکده پرستاری دانشگاه علوم پزشکی تهران (۱۳۸۵) ۴۸/۶٪ سوالات در تاکسونومی I، طراحی شده بودند<sup>۱۹</sup> کمبود استفاده از سوالات در سطوح تاکسونومی II و III نه تنها در آزمون های پرستاری بلکه در سوالات دروس تخصصی نظری مامایی طی دو نیم سال تحصیلی ۸۶-۱۳۸۵ نیز گزارش شده است<sup>۲۳</sup>. این مشکل نه تنها در علوم پایه بلکه در علوم بالینی پزشکی نیز دیده می شود. نمونه بارز آن که در مطالعات برخی از دانشگاه های علوم پزشکی کشور انجام گرفته است، پایین بودن درصد تاکسونومی II و III سوالات در آزمون های ارتقاء دستیاری است<sup>۱۲</sup>. البته رعایت استانداردهای تاکسونومی در حد استاندارد در طراحی آزمون های ارتقاء دستیاری سال ۱۳۸۸ در اساتید هیات علمی دانشگاه علوم پزشکی کرمانشاه نیز گزارش شده است<sup>۲۴، ۲۵</sup>. لزوم تحلیل کیفی سوالات برای ارتقا طراحی آزمون های چند گزینه ای از پیشنهادات پژوهش های مرتبط در این زمینه است<sup>۹، ۱۳</sup> که می تواند دارای نتایج مثبت در مدرسین رشته های مختلف علوم پزشکی باشد<sup>۸، ۱۲</sup>. و پیشنهاد شده است که با تدابیر اصلاحی توزیع تاکسونومی سوالات و اصول ساختاری ارتقاء یابند<sup>۱۳</sup>. نکته ای که باید به آن توجه داشت مطالعاتی است که ضریب تمییز بیشتری در سوالاتی

3. Newble DI, Cannon RA. A handbook for medical teachers. 4<sup>th</sup> ed. Boston: Kluwer academic publications 2001.
4. Briggs DC, Alonzo A, Schwab C, Wilson M. Diagnostic assessment with ordered multiple-choice items. Educational Assessment. 2006. 11(1), 33-63.
5. Oven A, Byrne G, ONill P. FDTL4: A core question bank linked to

learning outcomes for undergraduate medicine. Available at: <http://www.1tsn01.ac.uk/newsletter/01.3.htm1/corequestionbank>. Accessed May 10. 2009

6. Rudner LM. Scoring and Classifying Examinees Using Measurement Decision Theory. *Practical assessment, Research and Evaluation*. 2009; 14(8):1-12.

7. Grak yaraghi M, Toghyanifar N, Esmaili A, Avizhg M, Mazaheri M. Evaluation of MGQ on clinical Training in Esfahan Medical University 2005 – 2006. *Strides in Development of Medical Education*. 8<sup>th</sup> National Congress of Education – March 2007: 73 [Persian].

8. Javadi M, Abbaszade A, Borhani F, Abdoli R. Analytic Evaluation of MCQ in End Term Exams in Nursing School of Baam Strides in Development of Medical Education. 8<sup>th</sup> National Congress of Medical Education – March 2007: 217 [Persian].

9. Drakhsan F, Jahanihashemi H. Evaluation of End Term Exams of Ghazvin Medical University in 2007-2008 Special Supplement. 10<sup>th</sup> National Congress of Medical Education – May 2009: 308-309 [Persian].

10. Latifi SM, Shakornia A. Comparison of Experts behaviors and Views about MCQ Design in Ahvaz Medical University. *Strides in Development of Medical Education*. 8<sup>th</sup> National Congress of Medical Education – March 2007: 159-160 [Persian].

11. DoriPour F, Karimi Sh, PakAin J. Evaluation of Assay multiple choice Questions in Nursing Health Subjects. in 2006. *Strides in Development of Medical Education*. 8<sup>th</sup> National Congress of Medical Education – March 2007: 67 [Persian].

12. Haghshenas M, Vahid Shahi K, Mahmudi M, Shahbazenejad 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 in Development of Medical Education*. 2008; 5(2):120-127 [Persian].

13. Gholami Vafamehr V, Dadgostornia M. Assessment of Effect of Proficiency on MCQ in Reliability of clinical Medicine Query in Esfahan Medical university from 2002-2005. *Strides in Development of Medical Education*. 8<sup>th</sup> National Congress of Medical Education – March 2007:149 [Persian].

14. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The Levels of Difficulty and Discrimination Indices in Type A Multiple Choice Questions Of Pre-clinical Semester 1, Multidisciplinary Summative Tests. *IeJSME* 2009; 3 (1): 2-7.

15. Quadrelli S, Davoudi M, Galíndez F, Colt HG. Reliability of a 25-item low-stakes multiple-choice assessment of bronchoscopic knowledge. *Chest*. 2009; 135(2):256-8.

16. Vafamehr V, Dadgostornia M. Reviewing the results of qualitative and quantitative analysis of MCQs in Introduction to clinical medicine course. *IJME*. 2010; 10 (5):1146-1152 [Persian].

17. Kaveh Tabatabaee MS, Bahreiny Toosi MH, Derakhshan A, Khajeh Dalloee M, Gholami H. Analytic assessment of multiple-choice tests. *Journal of Medical Education*. 2003; 3(2):87-91.

18. Bachhoff 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.

19. Shaban M, Ramazani Badr F. Effect of Analysis of questions on Quality of Design Exams. 2007; *Hayat*. 13 (1): 5-15 [Persian].

20. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a Para-clinical multidisciplinary paper. *Ann Acad Med Singapore*. 2006; 35(2):67-7.



21. Zolfagari B, Adibi N, Drakhshonfor S, Tarsaz M, Karbasi A, Niromand D. Academic Achievement Tests. 1<sup>st</sup> ed. Esfahan. Neshat. 2001; 20,41,43,61 [Persian].
22. Mohagheghi MA, Vahid shahi K, Shakeri S, Sabori M, Razavi SM, Mohammadi M, Mirzai M. Evaluation of Some Quality index of Multiple choice Questions in the board exams in Medicine from 2005-2007. (Cited Sep, 2011). available at: [Http: //cgme . behdasht .gov.ir](http://cgme.behdasht.gov.ir).
23. Kazemi Ashraf, Ehsanpour Soheila, Hassanzadeh Akbar, Investigating the Academic Achievement Evaluation of Specialized Theoretical Courses of Midwifery BS Iranian Journal of Medical Education, 2010; 9(4):346-355 [Persian].
24. Roshanpour F, Pourmirza Kalhori R, Rezai M. Quantitative Analysis of Medical Promotion Exams in Kerman Shah Health and Education University in 2008. Special Supplement for 10th National Congress of Medical Education -May 2009; 335-336 [Persian].
25. Roshan Pour F, Pourmirza Kalhori R, Rezai M. Qualities Analysis of Medical Promotion Exams in Kermanshah health and Education University in 2008. Special Supplement for 10th National Congress of Medical Education -May 2009:337-338 [Persian]
26. Pourmirza Kalhori R, Razazian N, Najafi F, Almasi A, Hassanpour Dehkordi A. Educational purpose and medical students achievement in clinical training in an university hospital in kermanshak. Journal of Kermanshah University of Medical Science. 2010; 13(4). 344-352 [Persian].
27. Kibble JD, Johnson T. Are faculty predictions or item taxonomies useful for estimating the outcome of multiple-choice examinations? Adv Physiol Educ. 2011; 35(4):396-401
28. Khoshrang H, Taheri M, Asadi A, Heidarzadeh A. Quality of Residents' Promotion Exams before & after Educational Intervention in 2010-2011. Iranian Journal of Medical Education. 2013; 13(7): 559-560 [Persian].

## Analytic assessment of multiple choice questions (MCQ) of faculties' members' scholarship in Kermanshah University of Medical Sciences, 2010-2011

Reza Pourmirza Kalhori<sup>1</sup>,  
Fatemeh Darabi<sup>2</sup>,  
Farah Roshanpour<sup>\*3</sup>,  
Mansour Rezaei<sup>4</sup>

1. Department of Emergency Medicine, School of Paramedical, Kermanshah University of Medical Sciences, Kermanshah, Iran.

2. Department of Anesthesia, School of Paramedical, Kermanshah University of Medical Sciences, Kermanshah, Iran.

3. Vice Chancellery For Research And Technology, Kermanshah University of Medical Sciences, Kermanshah, Iran.

4. Department of Biostatistics and Epidemiology School of Health Kermanshah University of Medical Sciences, Kermanshah, Iran.

*\*Corresponding Author: Deputy of research and technology in Kermanshah University Medical Sciences, Kermanshah University of Medical Sciences, Kermanshah, Iran*

*E-mail: roshanpour\_f@yahoo.com*

### Abstract

**Introduction:** Multiple choice questions (MCQs) are a way to evaluate the learning in medical sciences. The purpose of this study was to analyze of MCQs exams in faculty members' scholarship in Kermanshah University of Medical Sciences in 2010 - 2011.

**Methods:** Data of this descriptive analytic study were composed of 156 MCQs exams that designed faculty members' scholarship. Quantity data was collected by discrimination index, difficulty level, standard deviation and reliability of test index (KR20) and the quality data by percent criterion of questions without structural defect and percent of Taxonomy I & II & III questions. Data was analyzed by SPSS-18 software and independent T test and Chi-square, ANOVA & Tucky.

**Results:** Mean reliability coefficient (KR20) in total exams was 0.68(±0.31), Mean difficulty level %0.56 (±0.15), Mean discrimination index 0.21(±0.15). Mean of questions without structural defect 73.88(±14.88), Mean of Taxonomy I 34.65(±15.78), Taxonomy II 41.34 (±13.14) and Taxonomy III 23.99 (±15.40). Teuky Test was shown statistically difference between reliability coefficient (KR20) in all faculties and between difficulty level in Paramedics Faculty and others (P=0.001).

**Conclusion:** The study showed that in discrimination index, Taxonomy II & III and questions without structural defect those designed by board members' scholarship in Kermanshah Medical University were not acceptable in standard level and suggested educated them for obtain promotion in design exams.

**Key words:** Multiple choice questions, reliability, difficulty index, discrimination index.