پژوهشی اصیل

کیفیت طراحی و تحلیل آزمون‌های پیش کارورزی دانشجویان پزشکی دانشگاه علوم پزشکی همدان در سال‌های ۱۳۹۷تا ۱۳۹۹

معصومه رستمی معز [۱]، مهدی بیگلرخانی [۲]، اعظم معیاری [۲*]

[۱] مرکز تحقیقات علوم بهداشتی، دانشگاه علوم پزشکی همدان، همدان، ایران.
[۲] گروه طب ایرانی، دانشکده پزشکی دانشگاه علوم پزشکی همدان، ایران.

مشخصات مقاله

تاریخچه مقاله:
دریافت: ۱٤۰۰/۰۸/۲۸
پذیرش: ۱٤۰۰/۱۲/۲۱
چاپ: ۱٤۰۰/۱۲/۲۸

نویسنده مسوول:
اعظم معیاری، گروه طب ایرانی، دانشکده پزشکی دانشگاه علوم پزشکی همدان، ایران.
ایمیل: a.meyari@umsha.ac.ir

چکیده

زمینه و هدف: ارزشیابی گام اصلی در فرایند آموزش است که توانایی فراگیران را می¬سنجد و تحلیل آزمونها به عنوان یک ابزار، میزان کارایی آن را مورد سنجش قرار می¬دهد. آزمون پیش کارورزی آزمونی است که صلاحیت دانشجویان پزشکی قبل از ورود به مرحله کارورزی را می¬سنجد. بنابراین تحلیل آن از اهمیت ویژه¬ای برخوردار است. لذا در این مطالعه به بررسی شاخص های کمی و کیفی آزمونهای پیش کارورزی دانشجویان پزشکی دانشگاه علوم پزشکی همدان در سالهای۹۷ تا ۹۹ پرداخته شد.

روش بررسی: در این مطالعه اسنادی کلیه سوالات پیش کارورزی شهریور ماه ۹۷ تا ۹۹ از نظر تاکسونومی سوالات، اشکالات ساختاری و ضرایب دشواری و تمیز بررسی شد. تاکسونومی سؤالات بر اساس تقسیم بندی بلوم، ساختار سؤالات بر اساس چک لیست میلمن و ضریب دشواری وتمایز سؤالات بر اساس فرمول های علمی با استفاده از فرم های اکسل موجود در دانشگاه مورد بررسی قرار گرفت. تجزیه و تحلیل داده ها با استفاده از نرم افزار spss20 انجام شد.

یافته‌ها: در سال ۹۷ ، ٥٦٪ سؤالات و در سال ۹۸ و ۹۹، ٦٤٪ سؤالات در سطح تاکسونومی ۲ و ۳ طراحی شده اند و بیش از ۹۰٪ سؤالات بدون ایراد ساختاری بودند. ضریب دشواری سوالات بطور میانگین در سطح مناسب ، اما ضریب تمایز سؤالات در سالهای ۹۷ و ۹۹ متوسط و در سال ۹۸ ضعیف بود.

نتیجه‌گیری: با توجه به تعداد زیاد سؤالات با ضریب تمایز منفی، توصیه می شود سؤالات، گزینه ها و کلید پاسخ های صحیح پس از طراحی آزمون، مورد بازبینی قرارگیرد.

واژگان کلیدی: آزمون پیش کارورزی، تجزیه و تحلیل آزمون، تاکسونومی، ساختار سؤال، سؤالات چند گزینه ای

مقدمه

ارزیابی یک گام اصلی در فرایند آموزش است که عملکرد تحصیلی فراگیران را در طی یک دوره‌ی آموزشی می‌سنجد. یکی از روش‌های ارزشیابی آموزش استفاده از سوالات چند گزینه‌ای است. تحلیل آزمون‌ها در واقع می‌تواند به عنوان یک ابزار آموزشی تعریف شود که صلاحیت آموزش فراگیران و همچنین شکاف بین اهداف آموزشی و میزان یادگیری را آشکار می‌سازد. روش‌های متعددی برای ارزیابی سطح شناخت یادگیرنده و اطمینان از دستیابی به اهداف آموزشی وجود دارد (۱).

سوالات چند گزینه‌ای یکی از انواع سوالات نوشتاری است که معمولا در فعالیت‌های دوره‌ای مبتنی بر محتوای نظری دوره درسی استفاده می‌شود (۲-٤). اگرچه سوالات چند گزینه‌ای به طور معمول سطح پایین دانش را ارزیابی می‌کنند، اما چنانچه به درستی طراحی شده و منطبق با سنجش اهداف آموزشی بکار گرفته شوند، می‌توانند برای ارزیابی سطوح بالای دانش، یعنی درک، تجزیه و تحلیل، کاربرد و حل مسئله به کار روند (٥).

همچنین برای نوشتن سؤالات چند گزینه‌ای خوب، علاوه بر دانش به مهارت خوب نوشتن نیز نیازمند است (٦, ۷).

برای تکمیل فرآیند آموزش در طول دوره، لازم است که مطالعه و تجزیه و تحلیل کیفیت سوالات انجام شود. بنابراین، تجزیه و تحلیل آزمون یک جزء یکپارچه از ارزیابی دوره است. به عبارت دیگر، ارزیابی آزمون یک فرآیند پویا است که با هدف بهبود سوالات و تدریس انجام می‌شود (۷). ارزشیابی دانشجو، یکی از وظایف مهمی است که یک استاد با آن مواجه است اما کیفیت بسیاری از روش‌های سنجش و امتحانات کمتر از حد مطلوب است. چنانچه ارزشیابی بر اساس اصول و موازین علمی صورت گیرد می‌توان آن را مهم‌ترین رکن آموزش و مؤثرترین عامل برای بهبود کیفیت یادگیری به شمار آورد (۸).

نظریه‌ی ارزیابی آزمون کلاسیک مبتنی بر معادله‌ای است که در آن نمره مشاهده شده به طور فرضی از یک نمره‌ی حقیقی و یک نمره خطا تشکیل شده است. در این مدل نمره‌ی حقیقی هر فرد ثابت است و

خطاهای اندازه گیری متغیر تصادفی خواهد بود. در نظریه‌ی کلاسیک، شاخص‌هایی همانند دشواری، میانگین، واریانس، اعتبار و روایی آزمون برآورد و محاسبه می‌گردد (۹). همچنین تجزیه و تحلیل سؤالات به دو روش کیفی و کمی انجام می‌گیرد. در تجزیه و تحلیل کمی سؤالات سه مولفه سطح دشواری یا آسانی، قدرت تمیز دادن سوال و گزینه‌های انحرافی بررسی می‌گردد (۱۰). بررسی ساختار سوالات با استفاده از چک لیست میلمن در بررسی کیفی سوالات می‌گنجد و به درصد سوالات بدون هیچ مشکلی ساختاری اشاره دارد (۳, ٤). مهم‌ترین اهداف میلمن در طراحی سوالات چند گزینه‌ای شامل قرار دادن بیشتر اطلاعات پرسش در تنه سوال، استفاده از سوال برای سنجش یک هدف یادگیری، استفاده از کلمات روان و واضح، عدم استفاده از کلمات یا عبارات منفی در گزینه-های سوال ویا برجسته کردن آن‌ها، عدم استفاده از پاسخ‌هایی با کلمات "همه گزینه‌ها" یا "هیچکدام"، عدم استفاده از گزینه‌های متناقض، استقلال سوالات از یکدیگر، معادل بودن سوالات در طول و ساختارهای واژگان، عدم استفاده از کلمات تکراری، عدم اشتباهات املایی و چینش مزتب گزینه‌ها است (۱۱).

علاوه بر این تجزیه و تحلیل کیفی، اشاره به درصد سوالاتی دارد که مبتنی بر هدف آموزشی با استفاده از تاکسونومی بلوم طبقه بندی شده است و شامل طبقه بندی سطح دانش (I)، درک و مفهوم (II)و کاربرد (III) است (۳, ۱۲). استفاده از سوالات چند گزینه‌ای معتبر و در سطوح تاکسونومی بالا به دانشجویان پزشکی کمک می‌کند تا درک مفهومی خود را افزایش دهند و برای پیچیدگی‌های تمرین بالینی با بهبود استدلال شناختی و توسعه دانش روشمند روبه رو شوند (۱۲).

آزمون پیش کارورزی آزمونی است که توسط آن صلاحیت دانشجویان رشته‌ی پزشکی قبل از ورود به مرحله کارورزی سنجیده می‌شود تا در صورت احراز شرایط لازم وارد مرحله بعد شوند که در واقع مرحله تصمیم گیری برای بیماران است. با توجه به اهمیت ویژه این آزمون به منظور سنجش دانش پزشکان آینده و در نهایت ارتقاء سلامت جامعه، واضح است که هر چه سوالات استانداردتر و بر مبنای علمی صحیح طراحی شده باشد ابزار بهتری برای افتراق دانشجویان فاقد صلاحیت علمی از دانشجویان واجد شایستگی علمی خواهد بود. در مورد موفقیت یا شکست در آزمون‌های پیش کارورزی، مطالعات متعددی منتشر شده و عوامل مختلفی بر نتایج آن تأثیرگذار گزارش شده است. موفقیت یا عدم موفقیت دانشجویان در رشته‌ی پزشکی عواملی مانند معدل دیپلم و معدل علوم پایه، فیزیوپاتولوژی و کارآموزی، نمره‌ی آزمون‌های پیش کارورزی، سابقه‌ی مشروط بودن، سهمیه پذیرش، جنسیت، مدت تحصیل، اشتغال دانشجو و بومی بودن مطرح و بررسی شده است که نتایج آن در برخی موارد متناقض است (۱۳). ارزیابی کلی آزمون‌های سوالات چند گزینه‌ای پیش کارورزی به روش علمی کمک می‌کند تا سطح آزمون‌ها مبتنی بر شواهد علمی ارتقا یابد و در نتیجه از بهبود آزمون‌ها برای ارتقا سطح آموزشی و ارزیابی یادگیری دانشجویان استفاده گردد (۱٤, ۱۵). بنابراین، با توجه به اهمیت آزمون‌های پیش کارورزی در ابتدای دوره‌ی کارورزی و نیاز به تحلیل آن (۱۳)، این مطالعه بمنظور ارزیابی علمی آزمون‌های سوالات چند گزینه‌ای پیش کارورزی دانشجویان پزشکی دانشگاه علوم پزشکی همدان در سه دوره اخیر (سال‌های ۹۷ تا ۹۹) انجام شد تا از نتایج آن برای ارتقا سطح آموزشی و ارزیابی یادگیری دانشجویان استفاده گردد.

روش بررسی

این مطالعه یک تحقیق اسنادی است. هدف پژوهش حاضر بررسی آزمون‌های پیش کارورزی رشته‌ی پزشکی در سال‌های ۹۷ تا ۹۹ می‌باشد. ابزارسنجش، با بررسی آزمون چند گزینه‌ای پیش کارورزی برگزار شده در مرکز آزمون‌های دانشگاه علوم پزشکی همدان (زیر مجموعه کلان منطقه آمایشی ۳ کشوری) انجام شده است.

در این مطالعه کلیه‌ی آزمون‌های پیش کارورزی رشته‌ی پزشکی برگزار شده در مرکز آزمون‌های دانشگاه علوم پزشکی همدان بر اساس نظریه تستی کلاسیک (۹) و چک لیست میلمن (۱۱) در سه دوره‌ی اخیر(شهریور۹۷ ، شهریور ۹۸، و شهریور ۹۹) مورد تجزیه و تحلیل قرار گرفت. آزمون‌ها از لحاظ شاخص‌های کمی (ضریب دشواری، ضریب تمیز و اعتبار آزمون) و شاخص‌های کیفی (درصد رعایت اصول طراحی ساختاری سوال و درصد طراحی سوالات در تاکسونومی سطح I،IIو III) مورد ارزیابی قرار گرفتند. در این مطالعه به منظور وجود چالش‌ها و اختلافاتی در افتراق و تعیین سطح تاکسونومی II و III ، سوالات بصورت سوالات با تاکسونومی بالا (II و III) و سوالات با تاکسونومی پایین(I) تقسیم بندی شد. مشکلات ساختاری سوالات با استفاده از چک لیست استاندارد میلمن (۱۱) سنجیده شد. این چک لیست حاوی ۱۸ سوال است و پاسخ آن بر اساس بله و خیر نمره دهی می‌گردد. هر سوال توسط پزشک متخصص و کارشناس در زمینه‌ی آموزش پزشکی بررسی شده و با توجه به چک لیست میلمن کلیه موارد موجود در چک لیست بررسی شده و در صورت عدم رعایت در چک لیست مربوط به هرسوال یادداشت گردید. همچنین در پایان تاکسونومی سوال توسط همان فرد کارشناس در زمینه‌ی آموزش پزشکی مشخص شد و در انتهای چک لیست یادداشت گردید. جهت برقراری ایجاد اتفاق نظر بین محققین (inter- rater reliability)، در طی جلساتی هماهنگی لازم برای نیل به این اهداف ایجاد شد.

در شاخص‌های کمی سطح دشواری بیانگر درصد افرادی است که به سؤال پاسخ درست داده‌اند، و از طریق تقسیم تعداد افرادی که به آزمون جواب صحیح داده‌اند به تعداد افرادی که در آزمون شرکت کرده‌اند، محاسبه می‌شود و کمتر از ۰/۳ آزمون دشوار و بین ۰/۳ تا ۰/۷ مناسب و بیشتر از ۰/۷ آزمون آسان دسته بندی می‌شود (۳). قدرت تمیز یا تشخیص، نشان دهنده‌ی قدرت تشخیص افراد با توانایی بالا (نمره‌ی بالا) و توانایی پایین (نمره‌ی پایین) است. دامنه‌ی قدرت تمیز سوال در محدوده ±۱ می‌باشد. محدوده صفر به مفهوم آن است که سؤال فاقد قدرت تشخیص است.محدوده‌ی کمتر از صفر، ۰/۱۲-۰ ، ۰/۳۹- ۰/۱۳و ۰/٤۰ و بیشتر به ترتیب نشاندهنده‌ی ضریب تمایز غیر قابل قبول، ضعیف، متوسط و خوب است (۳). این شاخص‌های کمی نظریه تستی کلاسیک (ضریب دشواری و ضریب تمیز) با استفاده از فرم‌های اکسل موجود در مرکز ازمون به ازای هر آزمون به دست آمد .

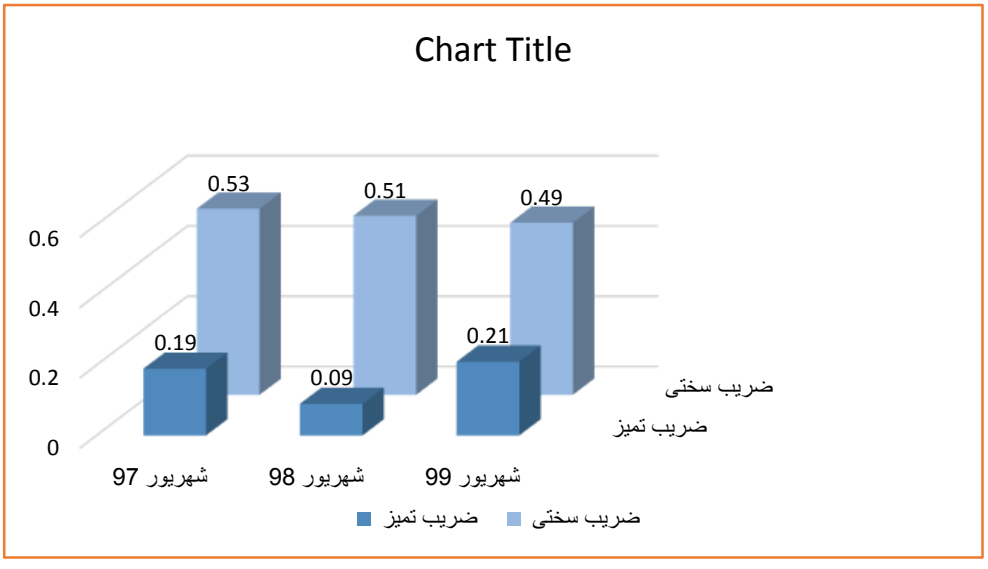
این مطالعه در کمیته‌ی اخلاق شورای پژوهشی دانشگاه علوم پزشکی همدان به شماره IR.UMSHA.REC.1398.740 مصوب شد. با توجه به اینکه از نمونه‌های انسانی استفاده نشد و داده‌ها از فرم‌های موجود در مرکز آزمون استخراج شد نیاز به رضایت نامه فردی نبود. داده‌ها با استفاده از نرم افزار spss ورژن ۲۰ مورد بررسی قرار گرفت. برای گزارش متغیرهای کمی با توجه به توزیع نرمال داده‌ها از میانگین و انحراف معیار استفاده شد. برای گزارش متغیرهای کیفی نیز نسبت‌ها و درصد بیان شد.

یافته‌ها

کلیه‌ی ۲۰۰ سوال سه دوره‌ی آخر برگزاری آزمون در سال‌های ۹۷،۹۸، و ۹۹ مورد بررسی قرار گرفت. این سوالات در ۱۷ حیطه‌ی مختلف (توسط ۱۷ گروه آموزشی) طراحی شده بود. تعداد شرکت کنندگان به ترتیب در سال‌های ۹۷، ۹۸ و۹۹ در دانشگاه علوم پزشکی همدان ۲٤ نفر، ۳۵ نفر و ۸٦ نفر بودند. میانگین اعتبار آزمون‌ها در سال‌های ۹۷ تا ۹۹ به ترتیب برابر با ۰/۸٦، ۰/۹۱، و ۰/۸۸ بود . هر سه آزمون از اعتبار خوبی برخوردار بودند. میانگین ضریب تمیز در سال‌های ۹۷ تا ۹۹ به ترتیب برابر با ۰/۱۹، ۰/۰۹، و ۰/۲۱ بود که نشان می‌دهد ضریب تمایز در سال-های ۹۷ و ۹۹ متوسط و در سال ۹۸ ضعیف بوده است (جدول ۱).

میانگین ضریب سختی در سال‌های ۹۷، ۹۸و۹۹ به ترتیب برابر با ۰/۵۳، ۰/۵۱، و ۰/٤۹ بود که نشان می‌دهد سوالات در هر سه سال از نظر دشواری، در سطح مناسبی قرار داشته است (نمودار ۱). تعداد سوالات آسان در آزمون‌های پیش کارورزی پزشکی در سال‌های ۹۷، ۹۸و ۹۹ به ترتیب ۵۵، ٤٦ و ۳۵ بوده است که هر سال از تعداد آن کاسته شده و به تعداد سوالات مناسب افزوده شده است (جدول ۱).

بیش از ۹۰٪ سوالات در سال‌های مورد بررسی بدون ایراد ساختاری بودند و در سال ۹۹ کمترین ایراد ساختاری در سوالات مشاهده شد. در سال ۹۷ تعداد ۱۱٦ سوال (٥٦٪ سوالات) و در سال‌های ۹۸ و ۹۹ تعداد ۱۲۷ سوال (٦٤٪ سوالات) در سطح تاکسونومی بالا (۲و۳) طراحی شده بود (جدول ۱).



شکل ۱: توزیع میانگین ضریب سختی و ضریب تمیز سوالات آزمون پیش کارورزی پزشکی در سال‌های ۹۷ تا ۹۹

جدول۱: بررسی توزیع فراوانی سوالات با تاکسونومی بالا، بدون ایراد ساختاری، دشواری و ضریب تمیز سوالات در آزمون‌های پیش کارورزی پزشکی در سال‌های ۹۷ تا ۹۹

متغیرها	شهریور ۹۷ (درصد) تعداد	شهریور ۹۸ (درصد) تعداد	شهریور ۹۹ (درصد) تعداد
سوالات با تاکسونومی بالا	۱۱٦(٪۵۸)	۱۲۷(٪٦٤)	۱۲۷(٪٦٤)
سوالات بدون ایراد ساختاری	۱۸۲(٪۹۱)	۱۸۱(٪۹۰)	۱۹۲(٪۹٦)
تعداد سوالات با قدرت تمیز منفی	٤۱ (۲۰/۵۰)	۳۹ (۱۹/۵۰)	۱۵ (۷/۵۰)
تعداد سوالات با قدرت تمیز ضعیف	۲٤ (۱۲)	۸٤ (٤۲)	۳۷ (۱۸/۵۰)
تعداد سوالات با قدرت تمیز متوسط	۱۰۰ (۵۰)	۷۵ (۳۷/۵۰)	۱۳۳ (٦٦/۵۰)
تعداد سوالات با قدرت تمیز خوب	۳۵ (۱۷/۵۰)	۲ (۱)	۱۵ (۷/۵۰)
تعداد سوالات آسان	۵۵ (۲۷/۵۰)	٤٦ (۲۳)	۳۵ (۱۷/۵)
تعداد سوالات مناسب	۱۰۷ (۵۳/۵)	۱۱۲ (۵٦)	۱۲۲ (٦۱)
تعداد سوالات سخت	۳۸ (۱۹)	٤۲ (۲۱)	٤۳ (۲۱/۵)

بحث

نتایج کلی این مطالعه نشان داد سطح دشواری آزمون‌ها در سال‌های مورد بررسی در سطح متوسط (حدود ۰/۵۰) بوده است. میانگین ضریب تمیز سوالات در سال‌های مورد بررسی در سطح متوسط و ضعیف (بین ۰/۰۹ تا ۰/۲۱) بود. نگاهی دقیق‌تر به نتایج نشان می‌دهد که بیش از ٦۰٪ سوالات در سال ۹۸ دارای ضریب تمایز منفی یا ضعیف بوده‌اند و در سال‌های ۹۷و ۹۹ این مقدار حدود ۳۲٪ و ۲۵٪ بوده است. با توجه به آن که ضریب تمایز هر آزمون هر چه به یک نزدیک‌تر باشد آزمون از قدرت تمیز بهتری برخوردار بوده و می‌تواند دانشجویان قوی و ضعیف در درس را به خوبی افتراق دهد بنابر این آزمون‌های مورد بحث بخصوص در سال ۹۸ به خوبی قادر به افتراق دانشجویان ضعیف از دانشجویان قوی نبوده‌اند. برای اینکه ضریب تمیز ارتقا یابد بایستی سطح دشواری سوالات مطلوب باشد زیرا سوالات دشوار و خیلی دشوار و همچنین سوالات بسیار آسان از قدرت تفکیک کمی برخوردار هستند (۹). در صورتی که سطح دشواری سوالات مطلوب باشد الگوی پاسخ‌ها و متن سوالات و گزینه‌های انحرافی باید مورد توجه قرار گیرد. سوالات با ضریب تمیز منفی نشان می‌دهد که در آن سؤال، گروه ضعیف بهتر از گروه قوی عمل کرده است. این گونه سؤال‌ها دارای ایرادات اساسی هستند که باید حذف شوند و یا در آن‌ها تجدید نظر اساسی صورت پذیرد. در واقع وجود گنگی و ابهام در تنه سوال و یا گزینه‌ها بسیار محتمل است . احتمال نادرست بودن گزینه کلید اعلام شده توسط طراح سوال یا اشتباه وارد شدن کلید پاسخ‌های صحیح در پاسخ نامه از دیگر دلایل آن می‌باشد (۳, ۹). نگاهی به مطالعات مختلف در طراحی سوالات آزمون‌های مشابه نشان‌دهنده‌ی شباهت‌ها و تفاوت‌هایی است. Nibret و همکاران در مطالعه‌ی خود در مورد آنالیز آزمون ۱۷٦ دانشجوی سال اول به نتایج مشابه مطالعه حاضر با ضریب تمیز ۰/۱٦ و ضریب سختی ۰/۵٦ بود (۱٦). در مطالعه‌ای که پورمیرزا کلهری و همکاران در سال ۱۳۹۲ درباره روند پنج ساله طراحی آزمون‌های ارتقاء دستیاری تخصصی پزشکی دانشگاه علوم پزشکی کرمانشاه انجام دادند ضریب دشواری کل سوالات ٦۲٪ و ضریب تمایز سوالات ۲۷٪ برآورد شده است که هر دو در محدوده مطلوبی بوده است (۱۷). در مطالعه‌ی عنبری در آزمون‌های ارتقاء دستیاری سال ۱۳۹۱، ٤٤/۷٪ از سوالات از نظر شاخص دشواری و ۵٤/۱٪ سوالات از نظر شاخص تمایز مناسب ارزیابی شده‌اند (۱۸). در مطالعه‌ی Khafagy و همکاران در مصر ضریب تمیز سوالات در سال ۲۰۱۳ نسبت به سال ۲۰۰۹ افزایش یافته بود و در سال ۲۰۱۳ به بالاتر از ۰/۳ رسیده بود و ضریب سختی سوالات از ٦۵ به ۵۵ کاهش یافته بود (۱۹). از آنجا که محتوای سوالات با هم متفاوت هستند نتایج متفاوت نیز قابل انتظار است اما به نظر می‌سد مقایسه روند روانسنجی آزمون‌ها در طی زمان از اهمیت بالاتری برخوردار باشد.

همچنین نتایج این مطالعه نشان می‌دهد که حدود ۵۸ ٪ سوالات در سال ۹۷ با تاکسونومی سطح بالا طراحی شده بود که این مقدار در سال‌های بعد به ٦٤٪ رسیده است. همچنین در هر سه آزمون، طراحی اغلب سوالات از نظر ساختاری بدون ایراد بودند. مقایسه این نتایج با مطالعات مشابه داخلی و یا خارجی نشان می‌دهد که این آزمون‌ها که مربوط به بدو ورود

دانشجویان به مرحله کارورزی و مسئولیت پذیری دانشجویان در قبال بیماران می‌باشد از نظر سطح تاکسونومی در سطح قابل قبولی طراحی شده‌اند. مطالعه‌ی صبوری و همکاران در سال ۱۳۹۸ در بررسی ۵ ساله سوالات بورد تخصصی رشته‌های مرتبط با سرطان بالغین از سال ۹۲ تا ۹۶ نشان داد که ٥٤٪ سوالات در سطح تاکسونومی ۱ و ۲۳٪ سوالات در سطح تاکسونومی ۲ طراحی شده بود (۲۰).در مطالعه‌ی پورمیرزا کلهری و همکاران بطور متوسط ۳۳/٤٪ سوالات با سطح تاکسونومی ۱ و ٦٦/٦٪ سوالات با سطح تاکسونومی ۲ و ۳ بوده و ٦۲/٦٪ سوالات بدون اشکال ساختاری طراحی شده بود (۱۷). مطالعه‌ی شکورنیا در مورد سوالات آزمون ارتقاء دستیاری در جندی شاپور در سال ۱۳۸۹ نشان داد که فقط ۷٪ سوالات در سطح تاکسونومی ۲ و ۳ طراحی شده است (۲۱).

مطالعه‌ی عنبری و همکاران در مورد ارزیابی آزمون ارتقاء دستیاران گروه-های بالینی دانشگاه علوم پزشکی اراک در سال ۹۱ نشان داد که ۸۰٪ سوالات بدون اشکال ساختاری و حدود ٦٥/۲٪ سوالات در سطح تاکسونومی ۲ و ۳ طراحی شده است (۱۸). در مطالعه‌ی بقایی و همکاران که میزان رعایت شاخص های سوالات چندگزینه‌ای در دانشکده‌ی پرستاری و مامایی دانشگاه ارومیه در سال ۹٤ را بررسی نمود، نتایج نشان داد ۸۵ درصد سوالات در سطح یک تاکسونومی بلوم بوده و در کل شاخص‌های تاکسونومی بلوم سوالات در سطح مطلوبی نبوده است (۲۲). مقایسه نتایج این مطالعه با مطالعات دیگر نشان می‌دهد که طراحی سوالات از نظر ایرادات ساختاری و تاکسونومی سوالات در سطح خوب و قابل قبولی طراحی شده است اما به نظر می‌رسد طراحان سوال بهتر است علاوه بر توجه به تاکسونومی سوالات و صحیح بودن ساختار سوالات و سطح دشواری آزمون‌ها، از نظر انتخاب گزینه‌های انحرافی مناسب نیز آموزش‌های لازم را گذرانده و علاوه بر آن پس از طراحی آزمون، سوالات ترجیحا توسط همکاران خبره و آموزش دیده مورد بازنگری قرار گیرند تا تنه سوال و گزینه‌های انحرافی مناسب‌تر انتخاب شده و کلید سوالات مورد بازنگری قرار گیرد. چنانچه معیاری و همکاران در سال۱۳۹۱ در مطالعه‌ی خود نشان دادند حدود ۳۵ درصد سؤالات ارتقاء دستیاری دارای اشکال ساختاری بوده که با مداخله انجام شده این میزان بطور معناداری کاهش یافته بود و همچنین میزان سوالات دارای تاکسونومی بالا از ۳۸٪ به ۵۳٪ افزایش یافته بود (۲۳).

نتایج این مطالعه می‌تواند برای بهبود یا ارتقا کمیت و کیفیت سوالات پیش کارورزی دانشجویان پزشکی دانشگاه‌های علوم پزشکی استفاده شود. دانش اساتید در ارزیابی و ارزشیابی یک فعالیت پویا و مداوم است. بنابراین پیشنهاد می‌شود که اساتید یا مربیان در کارگاه‌های حول محور تحولات آزمون و روانسنجی آزمون‌ها شرکت نمایند و دانش خود را به روز نمایند. همچنین به منظور به حداقل رساندن سوالات با ضریب تمایز منفی قبل از برگزاری آزمون، سوالات توسط همکاران گروه مورد بازبینی قرار گیرد تا اشکالات آن‌ها اصلاح شود. از آنجا که بیشتر آزمون‌های ساخته شده در دانشکده‌ها و توسط اساتید از نوع عینی هستند، تجزیه و تحلیل آزمون‌ها یا حداقل برخی از آن‌ها برای مدرسان توصیه می‌شود. همچنین پیشنهاد می‌گردد با توجه به وجود سیستم الکترونیک آزمون‌ها و امکان استخراج تجزیه و تحلیل آزمون‌ها یک واحد خاص مدیریت آزمون در نظر گرفته شود که مسئول تجزیه و تحلیل سوالات و مسئول بازخورد به اساتید و حتی راهنمایی و مشاوره در خصوص ارتقا طراحی و تاکسونومی سوالات آزمون‌ها باشد.

نتیجه گیری

نتایج این مطالعه نشان داد که سوالات آزمون پره انترنی از سال ۱۳۹۷ تا ۱۳۹۹ از نظر شاخص‌های کیفی نسبت به سال‌های قبل دارای کیفیت مناسبتری بوده است. بطوری‌که سوالات با تاکسونومی بالاتری طراحی شده و طراحان سوال خود را ملزم به رعایت اصول صحیح ساختار سوالات کرده‌اند. اما از نظر شاخص‌های کمی وجود تعداد بالای سوالات دارای ضریب تمایز ضعیف و منفی، لزوم بازنگری مجدد را خاطر نشان می‌سازد. لذا توصیه می‌شود قبل از طراحی سوالات اقدام به برقراری جلسات آموزشی و ارائه‌ی بازخورد مناسب و بجا شده و پس از طراحی سوالات نیز، گزینه‌های انحرافی و پاسخ صحیح مجددا مورد بازبینی قرار گیرند.

تضاد منافع

پژوهشگر در هیچ یک از مراحل تحقیق تضاد منافعی نداشته است.

تقدیر و تشکر

نویسندگان این مقاله بدینوسیله کمال تشکر و قدردانی خود را از کارشناسان مرکز آزمون دانشگاه علوم پزشکی همدان که ضمن رعایت محرمانه بودن نتایج آزمون، نهایت همکاری را در مورد در اختیار گذاردن سوالات آزمون‌ها و شاخص‌های مورد نیاز داشتند، اظهار و اعلام می‌دارند.

References

1. Iqbal MZ, Khan RA, Razaq N. Assessment of non-functional distracters in multiple choice questions: a descriptive analysis. *Pak J Physiol*. 2016;12(2):47-9.

2. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *JPMA-J Pak Med. Assoc*. 2012;62(2):142-147.

3. Saif A. Educational measurement, assessment and evaluation. Tehran: Doran Publications. 2004;128.

4. Amin Z, Chong YS, Khoo HE. Practical guide to medical student assessment:World Scientific; 2006.

5. Roediger III HL, Marsh EJ. The positive and negative consequences of multiple-choice testing. *J Exp. Psychol: Learn. Mem. Cognit.* 2005;31(5):1155-59.

6. Talebi GA, Ghaffari R, Eskandarzadeh E, Oskouei AE.. Item Analysis an Effective Tool for Assessing Exam Quality, Designing Appropriate Exam and Determining Weakness in Teaching. *Res Dev Med Educ.* 2013; 2(2): 69-72.

7. Abbasi H, Falsafinejad MR, Delavar A, Farrokhi NA, Mohagheghi MA. The Comparison of Two Models for Evaluation of Pre-internship Comprehensive Test: Classical and Latent Trait. *Iran. J Med. Educ.* 2013;13(3):167-78.

8. Amin MM, Shayan S, Hashemi H, Poursafa P, Ebrahimi A. Analysis of multiple choice questions based on classical test theory. *Iran. J Med. Educ.* 2011;10(5):719-25(persian).

9. pourmirza kalhori R, rezeai M, Karami Matin B, Roshan Pour 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. *J Med. Educ. Develop*. 2014; 8 (4) :64-75

10. Haladyna T M, Downing S M, Rodrigues M C. A review of multiple choise item writing guidelines for classroom assessment. *Appl. Measurement Educ*. 2002; 15(3):309-334.

11. Zaidi NLB, Grob KL, Yang J, Santen SA, Monrad SU, Miller JM, et al. Theory, process, and validation evidence for a staff-driven medical education exam quality improvement process. *Med Sci Educ*. 2016;26(3):331-6.

12. Khazaei MR, Zarin A, Rezaei M, Khazaei M. Factors affecting the results of comprehensive pre-internship exam among medical students of Kermanshah University of Medical Sciences. *Korean j med. educ.* 2018;30(2):131-139.

13. Musa A, Shaheen S, Ahmed A. Distractor analysis of multiple choice questions: A descriptive study of physiology examinations at the Faculty of Medicine, University of Khartoum. *Khartoum Med. J.* 2018;11(1).

14. Jafari F, Hadavand M, Samadpour M, Azami FH, Behbahan SEB. Evaluation of Pre-Internship Comprehensiv Exams Scores and their Predictive Factors. *Biomed Pharmac J.* 2013;6(2):307-13.

15. Nibret TK. Post exam analysis: Implication for intervention. *bioRxiv*. 2019; 1:510081.

16. Anbari Z, Jadidi R. Assessment of the resident's promotion exam: One step to validity of competency measurement in Arak University of Medical Sciences.*J med. educ. develop.* 2013; 7 (4) :52-62

17. Khafagy G, Ahmed M, Saad N. Stepping up of MCQs' quality through a multi-stage reviewing process. *Educ Primary Care*. 2016;27(4):299-303.

18. Sabouri M, Arbabi F, Dehghani Poudeh M. The Investigation into the Board Examinations in Majors Related to Adult Cancer in Iran. *Iran. J Med. Educ.* 2020; 20 :272-278.

19. Shakurnia A, Mozaffari A, Khosravi Brougeni A. Survey on Structural of MCQs of residency exam in AJUMS. *Jundishapur Sci Med J.* 2010;8(4):492-502.

20. Baghaei R, Feizi A, Shams S, Naderi J, Rasouli D. Evaluation of the nursing students' final exam multiple-choice questions in Urmia University of medical sciences. *J Urmia Nurse. Midwifery Fac*. 2016;14(4291):291-9.

21. Meyari A, Beiglarkhani M, Zandi M, Vahedi M, Miresmaeili A. The Effect of Education on Improvement of Multiple Choice Questions' Designing in Annual Residency Exams of Dental School. . *Iran. J Med. Educ*. 2012;12(1):36-45.

Rostami-Moez M, Biglarkhani M, Meyari A. Quality of design and analysis of pre-internship exams for medical students of Hamadan University of Medical Sciences in the years 2018 to 2020. J Med Educ Dev. 2022; 14 (44):36-42