

مقایسه نتایج ارزیابی آزمون ارتقاء دستیاران در گروه‌های آموزش بالینی دانشگاه علوم پزشکی اراک

زهرة عنبری*^۱، رحمت اله جدیدی^۲

۱- استادیار، دکترای مدیریت بهداشت و درمان، دانشگاه علوم پزشکی اراک، اراک، ایران
۲- استادیار، دکترای مدیریت بهداشت و درمان، دانشگاه علوم پزشکی اراک، اراک، ایران

تاریخ پذیرش: ۱۳۹۱/۸/۲۷

تاریخ دریافت: ۱۳۹۱/۵/۳

چکیده

سابقه و اهداف: طراحی ابزاری مناسب برای سنجش توانمندی دستیاران بنا بر نقش ویژه آنان در نظام آموزشی درمانی دانشگاه‌ها بیش از پیش مورد توجه برنامه ریزان آموزش عالی قرار گرفته است. این پژوهش با هدف ارزیابی و مقایسه آزمون ارتقاء دستیاران در دانشگاه علوم پزشکی اراک انجام گرفت.

روش بررسی: در این پژوهش توصیفی - تحلیلی ۷۵۰ سؤال از آزمون ارتقاء دستیاران در گروه‌های جراحی، داخلی، اطفال، زنان و بیهوشی از ابعاد رعایت اصول ساختاری، سطح پوشش تاکسونومی، شاخص‌های تمیز و دشواری و نیز میزان پوشش اهداف یادگیری در دانشگاه علوم پزشکی اراک مقایسه شد. ابزار جمع‌آوری داده‌ها شامل: چک لیست استاندارد میلمن، چک لیست تعیین سطح تاکسونومی و پوشش اهداف و نیز دستگاه تصحیح اوراق (OMR) بود. روایی و پایایی ابزارها تأیید گردید و داده‌ها بر اساس آزمون‌های x^2 و anova تحلیل شد.

یافته‌ها: نتایج نشان داد، گروه زنان با میانگین $(4/5 \pm 4/2)$ دارای کم‌ترین اشکال ساختاری، گروه داخلی از بعد پوشش بالاترین سطح تاکسونومی (40% در سطح کاربرد و 47% در سطح فهم) و گروه جراحی بالاترین درصد پوشش اهداف یادگیری ($90/7\%$) و نیز به عنوان مناسب‌ترین سؤالات از بعد شاخص دشواری ($67/3\%$) و تمیز ($73/5\%$) سؤالات در مقایسه با سایر گروه‌ها ارزیابی شد. بین اشکالات ساختاری، پوشش اهداف یادگیری و سؤالات استاندارد در بین گروه‌های آموزش بالینی اختلاف معناداری مشاهده گردید ($P=0/001$).

نتیجه‌گیری: این پژوهش بر لزوم تقویت تیم‌های ارزیابی آزمون، ارتقاء سیستم‌های الکترونیکی تحلیل آزمون، برگزاری کارگاه‌های آموزشی و کنترل سؤالات قبل از برگزاری آزمون تأکید نموده و اعمال سیاست‌های تشویقی و انجام مشاوره با کارشناسان تعلیم و تربیت و روان سنجی آزمون را به عنوان اثربخش‌ترین استراتژی‌ها ذکر می‌نماید.

واژه‌های کلیدی: مقایسه، آزمون، ارتقاء دستیاران، دانشگاه علوم پزشکی

* (نویسنده مسئول)؛ تلفن: ۴۱۷۳۶۴۱-۰۸۶۱، آدرس الکترونیکی: dr.anbari@arakmu.ac.ir

ارجاع به این مقاله به صورت زیر است:

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

مقدمه

بخش مهمی از فرآیند آموزش، ارزیابی پیشرفت تحصیلی دانشجویان می‌باشد (۱) و چنانچه بر اساس اصول و موازین علمی و به طور مستمر انجام پذیرد، می‌توان آن را به عنوان مهم‌ترین رکن آموزش به شمار آورد (۲) در حقیقت، ارزیابی تحصیلی به اندازه‌گیری پیشرفت دانشجو بر حسب معیارهای تعریف شده و بر اساس دوره آموزشی مربوطه که دانشجو باید به استانداردهای مورد نظر دست یابد، تعریف می‌شود (۳) و ارزشیابی، انجام ارزیابی به منظور ارزش گذاری و قضاوت درباره کیفیت و اثربخشی یک برنامه، روش‌های یاددهی، رتبه بندی فراگیران و یا تأثیر یک دوره آموزشی محسوب می‌شود (۴-۳). انجام ارزشیابی موثر نه تنها باعث افزایش انگیزه در دانشجویان شده، بلکه مدرس را در ارزیابی فعالیت‌های خود یاری می‌دهد (۶) در آموزش پزشکی نیز برای تعیین سطح معلومات، مهارت، هوش و استعداد دانشجویان از وسیله‌ای به نام آزمون استفاده می‌شود (۷) بنابراین مطابق با این تعریف، روش‌های مورد استفاده در ارزیابی دانشجویان باید متناسب و سازگار با اهداف ویژه یادگیری باشد (۱) چرا که اهداف آموزشی تعیین کننده مجموعه آموخته‌های نظری و عملی دانشجویان محسوب می‌شود که یادگیرنده باید در پایان دوره آموزشی آن را کسب کرده باشد (۲). تحقیقات انجام شده توسط گرون لوند (Gronlund)، فلاحتی و اپستین (Epstein) نشان می‌دهد که علیرغم اهمیت طراحی آزمون‌های استاندارد، متأسفانه کیفیت بسیاری از روش‌های سنجش و ارزیابی دانشجویان، بسیار کمتر از حد مطلوب می‌باشد و لزوم توجه بیش از پیش را می‌طلبد (۸-۱۰) همچنین گرگ یراکی و همکاران نیز در تحقیق خود این عدم توجه به طراحی ابزار و روش‌های مناسب ارزیابی را، عاملی مهم و کلیدی در کاهش انگیزه دانشجویان ذکر نموده و آن را یکی از دلایل ناکامی سیستم آموزشی در تحقق اهداف مطلوب خود، بر می‌شمرند (۱۱). بنابراین ارزیابی از توانمندی دانشجویان گروه‌های پزشکی خصوصاً دستیاران، بنابر نقش ویژه آنان در نظام آموزشی درمانی دانشگاه‌ها و نیز نقش مهم این گروه از

دانش‌آموختگان در ارتقاء نظام سلامت، بسیار حائز اهمیت بوده و توجه به برنامه‌های ارتقاء توانمندی دستیاران و نیز استانداردهای آزمون‌ها با هدف سنجش دقیق از میزان توانمندی آنان مورد توجه برنامه ریزان آموزش پزشکی قرار گرفته است (۱۲) بنابراین تدوین شیوه ارزیابی صحیح در فرآیند آموزش این گروه از دانشجویان و نیز ارائه سؤالات استاندارد و منطبق با اصول و قواعد مربوط به طراحی سؤالات چند گزینه‌ای، اجتناب ناپذیر خواهد بود (۱۳). در حال حاضر از میان ارزیابی‌های تراکمی یا پایانی (Summative Assessment)، که در مقطع دستگیری انجام می‌شود، آزمون‌های چند گزینه‌ای (Multiple Choice Question)، رایج‌ترین آزمون‌ها در ارتقاء دستیاران محسوب می‌شود (۴، ۱۲). این نوع آزمون‌ها از درجه پایایی بالایی برخوردارند، اما مسئله روایی محتوا (Content Validity)، روایی ساختار (Construct Validity)، توزیع سطح دانش و رعایت اصول ساختاری آن‌ها همواره مورد توجه محققین بوده است (۱۴). این آزمون‌ها اگر به درستی طراحی شوند، اصولاً سه سطح دانش از حیطة شناختی (Cognitive Domain) هرم بلوم (Blum) یعنی سطوح یادآوری (Recall)، فهم (Comprehension) و کاربرد (Application) را می‌تواند ارزیابی نماید (۱۵-۱۷). بنابراین ارزیابی حیطة مهارتی (Psychomotor Domain) دانشجو مانند: توانمندی در انجام یک پروسیجر بالینی، توانایی در برقراری ارتباط مؤثر با بیماران و نیز ارزیابی قدرت خلاقیت او، از طریق آزمون‌های چند گزینه‌ای امکان پذیر نبوده و باید از سایر روش‌های ارزیابی متناسب با نوع مهارت مورد آزمون، استفاده شود (۴) بر این اساس می‌توان اذعان نمود که طراحی و آزمون‌سازی استاندارد، نیازمند دانش تخصصی است و مدرسین برای ارزیابی تدریس خود، تعیین پیشرفت تحصیلی دانشجویان و تعیین کیفیت یادگیری، نیازمند دانستن مفاهیم پایه خواهند بود (۱۸) اما اصولاً اساتید تمایل به طراحی سؤالاتی دارند که سطوح پایین یادگیری دانشجو مانند حافظه یا یادآوری را

ارزیابی نماید. چون طراحی چنین سؤالاتی نیازمند صرف زمان و دقت کمتری خواهد بود. گرچه نمرات دانشجویان در چنین آزمون‌هایی دارای تورش (Bias) بوده و کسب نمرات چنین آزمون‌هایی وابسته به عواملی چون توانایی دانشجو در تست‌زنی و یا توانایی وی در حذف کردن برخی از گزینه‌ها بدون داشتن دانش مربوطه بوده و ممکن است نمره آزمون دانشجو را افزون بر میزان واقعی برآورد نماید (۱۹). در طی دو سال گذشته نیز، آزمون‌های ارتقاء دستیاران به صورت متمرکز و در سطح ملی برگزار می‌شد که در حال حاضر به صورت درون دانشگاهی برگزار می‌شود. لذا ارتقاء سطح روایی محتوا، روایی ساختار، توزیع سطح دانش و رعایت اصول ساختاری سؤالات بسیار حائز اهمیت بوده و توجه به طراحی آزمون‌های استاندارد، نقش مهمی در تعیین میزان واقعی توانمندی دستیاران خواهد داشت. پژوهش حاضر نیز با هدف مقایسه نتایج ارزیابی آزمون ارتقاء دستیاران در گروه‌های آموزش بالینی در دانشگاه علوم پزشکی اراک به منظور تعیین نقاط قوت و ضعف آزمون ارتقاء در گروه‌های مختلف، انجام گردید.

روش بررسی

در این پژوهش توصیفی - تحلیلی، که در شهریور ماه سال ۱۳۹۰ در دانشگاه علوم پزشکی اراک انجام گرفت، تعداد ۷۵۰ سؤال مربوط به آزمون‌های چند گزینه‌ای ارتقاء دستیاران سال اول تا سوم دوره دستیاری، در گروه‌های: جراحی، داخلی، اطفال، زنان و بیهوشی و به تفکیک هر گروه شامل ۱۵۰ سؤال مورد ارزیابی و مقایسه قرار گرفت. ابعاد مورد بررسی در گروه‌های مختلف آموزشی دارای دستیار عبارت بودند از: الف) - مقایسه از بعد رعایت اصول ساختاری سؤالات چند گزینه‌ای میلمن (Millmen) ب) - مقایسه از بعد پوشش سطح تاکسونومی سؤالات در حیطه شناختی بلوم شامل: تاکسونومی I (سؤالات مبتنی بر حافظه و یادآوری)، تاکسونومی II (سؤال مبتنی بر فهم) و تاکسونومی III (سؤال مبتنی بر کاربرد مفاهیم) ج) - مقایسه از بعد تعیین شاخص دشواری (Difficulty Index) و شاخص تمیز (Discrimination Index) (سؤالات، د) - و بالاخره مقایسه گروه‌ها از بعد میزان پوشش

اهداف اصلی یادگیری (Core curriculum)، تحت عنوان سؤالات Must learn (سؤالاتی که یادگیری آن برای دانشجویان در اولویت اول می‌باشد) و نیز سؤالات None core curriculum یا Best learn یا سؤالاتی که یادگیری آن برای دانشجویان در اولویت اول نمی‌باشد. در این پژوهش، سؤالی استاندارد تلقی گردید که دارای چهار ویژگی شامل: پوشش بالاترین سطح تاکسونومی، فاقد اشکال ساختاری، ضرایب تمیز و دشواری مناسب و جزء core curriculum بود. ابزار جمع آوری داده‌ها به منظور بررسی و مقایسه سؤالات آزمون ارتقاء دستیاران از بعد رعایت اصول ساختاری، چک لیست استاندارد شده و ۱۲ سؤالی میلمن بود (۲۰، ۲۱) که ۷۵۰ سؤال به طور جداگانه طبق اصول میلمن و توسط تیمی از استادیاران مرکز مطالعات و توسعه آموزش پزشکی که نسبت به طراحی آزمون‌های چندگزینه‌ای مسلط و صاحب نظر بودند، مورد بررسی قرار گرفت. همچنین ابزار تعیین شاخص تمیز و دشواری سؤالات آزمون ارتقاء دستیاران دستگاه تصحیح اوراق با عنوان دستگاه OMR بود. ضریب دشواری توسط این سیستم از محاسبه مجموع تعداد جواب‌های صحیح گروه بالا با تعداد جواب‌های صحیح گروه پایین تقسیم بر مجموع تعداد نمرات گروه بالا بر تعداد نمرات گروه پایین $(P=UP+LP/U+L)$ به دست آمد و ضریب تمیز $(D=UP-LP/U)$ نیز از تفاضل تعداد جواب‌های صحیح گروه بالا بر تعداد جواب‌های گروه پایین تقسیم بر تعداد دانشجویان یک گروه محاسبه گردید. به منظور تعیین سطح تاکسونومی سؤالات و نیز میزان پوشش اهداف یادگیری از چک لیستی که به روش دلفی (Delphi technique) تکمیل گردید استفاده شد. این چک لیست توسط تیمی متشکل از مدیران گروه‌های آموزش بالینی دارای دستیار و نیز استادیاران مرکز مطالعات و توسعه آموزش پزشکی تکمیل گردید به طوری که سطح تاکسونومی سؤالات در حیطه شناختی بلوم و نیز میزان پوشش اهداف اصلی یادگیری (core curriculum) مطابق با کوریکولوم معاونت آموزشی وزارت متبوع در هر گروه آموزشی، توسط این تیم خبره برای هر سؤال ارزیابی شد. روایی و پایایی چک لیست

بررسی تفصیلی نشان داد: از میان ۱۷٪ سؤالات دارای اشکالات ساختاری در آزمون ارتقاء "گروه داخلی"، ۵/۳٪ مربوط به بکار بردن گزینه‌های ترکیبی (۸ سؤال) و ۴/۶٪ مربوط به عدم شفافیت لغات در ساقه و گزینه سؤالات بود (۷ سؤال) و ۴٪ سؤالات دارای پاسخ‌هایی بیش از یک گزینه بود. به طوری که در ساقه و گزینه‌ها نیز غلط‌های املایی مشاهده گردید (۶ سؤال). این تحقیق نیز نشان داد از ۱۷٪ اشکالات ساختاری در آزمون ارتقاء دستیاران "گروه جراحی"، بیشترین اشکالات به ترتیب شامل: ۹/۳٪ مربوط به عدم همسنگ بودن گزینه‌ها از نظر طول ساختار لغوی و نیز سبک نگارش (۱۴ سؤال)، ۴٪ مربوط به متضاد بودن گزینه‌ها با یکدیگر در پاسخ‌ها بود. از ۱۸٪ سؤالات دارای اشکالات ساختاری در آزمون ارتقاء دستیاران در گروه‌های آموزشی "زنان و بیهوشی"، بیشترین اشکالات ساختاری سؤالات مربوط به "عدم همسنگ بودن گزینه‌ها از نظر طول، ساختار لغوی و سبک نگارش (۶٪ سؤالات) و نیز در ۴٪ سؤالات این اشکالات مربوط به عدم تمایز سؤالات منفی در تنه سؤال از سایر کلمات "ارزیابی گردید و "گروه اطفال" در ۳۲/۶٪ از سؤالات خود (۴۹ سؤال)، تنه سؤالات منفی را از سایر کلمات متمایز ننموده بود و ۶٪ سؤالات (۶ سؤال)، دارای پاسخ‌های بیش از یک گزینه و نیز دارای گزینه‌های ترکیبی (همه موارد، هیچ کدام) بود.

این پژوهش همچنین نشان داد، گروه داخلی از نظر پوشش سطح تاکسونومی III (۴۰٪) و تاکسونومی II (۴۷٪)، در مقایسه با سایر گروه‌ها به عنوان بهترین سؤالات ارزیابی شد و گروه جراحی از نظر پوشش اهداف یادگیری مطابق با must learn در کوریکولوم آموزشی (۹۰/۷٪) و نیز به عنوان گروه ارائه دهنده استانداردترین سؤالات آزمون ارتقاء دستیاران (۷۶٪) تعیین گردید. در مجموع ۳۲/۸٪ از کل سؤالات آزمون ارتقاء دستیاران در پنج گروه مورد بررسی، در تاکسونومی III و ۳۲/۴٪ در تاکسونومی II قرار داشت و ۴۲٪ (۳۱۵ سؤال) از کل سؤالات (۷۵۰ سؤال)، مطابق با پوشش core curriculum و ۳۴٪ (۲۵۵ سؤال)، به عنوان سؤال استاندارد ارزیابی شد.

میلن نیز به لحاظ استاندارد بودن مورد تأیید قرار گرفت (۲۱،۲۰). روایی چک لیست مربوط به تعیین سطح تاکسونومی و پوشش اهداف یادگیری بر اساس نظرات اساتید مجرب و توانمند گروه آموزش پزشکی و نیز اساتید گروه‌های آموزش بالینی دارای دستیار، تأیید شد. پایایی این چک لیست نیز بر اساس میزان همخوانی نظرات دو ارزیابی کننده به طور هم‌زمان مورد بررسی قرار گرفت که پایایی چک لیست تعیین سطح تاکسونومی سؤالات به میزان $r=0/79$ و پایایی میزان پوشش اهداف یادگیری سؤالات، $r=0/81$ به میزان محاسبه گردید که مورد تأیید نهایی قرار گرفت. به لحاظ رعایت اصول اخلاقی در این پژوهش، از معاونت آموزش و تحقیقات دانشگاه مجوز کتبی اخذ گردید و هماهنگی‌های لازم با ریاست محترم دانشکده پزشکی مبنی بر اخذ نظرات اعضای محترم هیئت علمی گروه‌های آموزشی دارای دستیار، انجام شد. داده‌ها بر اساس نرم افزار epi و آزمون anova جهت مقایسه اختلاف میانگین بین اشکال ساختاری سؤالات و نیز درصد پوشش اهداف یادگیری و درصد سؤالات استاندارد با گروه‌های آموزشی استفاده شد و از "آزمون X^2 " جهت تعیین ارتباط یا استقلال بین دو متغیر کیفی (مانند تعیین ارتباط بین شاخص‌های تمیز و دشواری با گروه‌های آموزشی و نیز تعیین ارتباط بین درصد پوشش تاکسونومی با گروه‌های آموزشی) استفاده گردید.

یافته‌ها

این پژوهش نشان داد که سؤالات آزمون ارتقاء دستیاران در گروه آموزشی زنان با میانگین $(4/5 \pm 4/2)$ ، دارای کم‌ترین اشکال ساختاری و گروه اطفال با میانگین $(7/5 \pm 13/6)$ دارای بیشترین اشکال ساختاری بود (جدول ۱). لازم به ذکر است که بیشترین اشکال ساختاری در کل آزمون ارتقاء، مربوط به عدم تعیین لغات منفی در ساقه سؤال (۱۸٪) و نیز عدم همسنگ بودن گزینه‌ها از نظر طول و ساختار لغوی و نیز سبک نگارش سؤالات (۶/۷٪) ارزیابی گردید و در مجموع ۲۰/۳٪ از کل سؤالات آزمون ارتقاء دستیاران در پنج گروه مورد بررسی، دارای اشکال ساختاری بود.

به آزمون X^2 جهت سنجش ارتباط یا استقلال بین درصد پوشش تاکسونومی و گروه‌های آموزشی با میزان پذیرش $P < 0/05$ ، ارتباط معناداری ارزیابی شد ($P = 0/001$).

مطابق جدول ۲، با توجه به میزان پذیرش $P < 0/05$ ، بین شاخص تمیز و دشواری سؤالات نیز با گروه‌های آموزشی، ارتباط معناداری وجود داشت ($P = 0/001$).

همچنین $44/7\%$ کل سؤالات از نظر شاخص دشواری و $54/1\%$ سؤالات از نظر شاخص تمیز، مناسب ارزیابی شدند (جدول ۲). مطابق جدول ۱، با توجه به آزمون آنالیز واریانس یک طرفه anova، جهت مقایسه میانگین در گروه‌های آموزشی مختلف، اختلاف معناداری از لحاظ میانگین اشکال ساختاری و نیز درصد پوشش اهداف یادگیری و درصد سؤالات استاندارد با گروه‌های آموزشی وجود داشت ($P = 0/001$). همچنین با توجه

جدول ۱: مقایسه میانگین اشکالات ساختاری، سطح تاکسونومی، میزان پوشش اهداف یادگیری و درصد سؤالات استاندارد در آزمون ارتقاء

دستیاران دانشگاه علوم پزشکی اراک در سال ۱۳۹۰

گروه آموزشی دارای دستیار	اشکالات ساختاری سؤالات (اصول میلمن) *	درصد پوشش تاکسونومی *			درصد پوشش اهداف یادگیری * (core curriculum)	+درصد سؤالات استاندارد در گروه‌ها *
		I	II	III		
گروه جراحی	درصد (تعداد) (میانگین \pm انحراف معیار) $(2/8 \pm 3/4)$	23% (۳۴)	40% (۶۰)	37% (۵۶)	$90/7\%$ (۱۳۶)	76% (۱۱۴)
گروه داخلی	$(2/7 \pm 1/7)$	13% (۲۰)	47% (۷۰)	40% (۶۰)	$57/3\%$ (۸۶)	52% (۷۸)
گروه اطفال	$(7/5 \pm 13/6)$	51% (۷۶)	11% (۱۷)	38% (۵۷)	20% (۳۰)	$17/3\%$ (۲۶)
گروه بیهوشی	$(3/6 \pm 2/9)$	46% (۶۹)	33% (۴۹)	21% (۳۲)	$24/7\%$ (۳۷)	$12/7\%$ (۱۹)
گروه زنان	$(4/5 \pm 4/2)$	41% (۶۲)	31% (۴۶)	28% (۴۲)	$32/7\%$ (۴۹)	12% (۱۸)
جمع از ۷۵۰ سؤال	$(4/22 \pm 6/45)$	$34/8\%$ (۲۶۱)	$32/4\%$	$32/8\%$	42% (۳۱۵)	34% (۲۵۵)
			(243)	(246)		(152)

* P-Value = $0/001$

+ منظور سؤالاتی است که فاقد اشکال ساختاری و در سطح تاکسونومی دو (فهم) یا سه (کاربرد) ازهرم بلوم و نیز جزء بایدهای یادگیری بودند.

جدول ۲: مقایسه توزیع فراوانی نسبی ضریب دشواری و تمیز سؤالات آزمون ارتقاء دستیاران دانشگاه علوم پزشکی اراک در سال ۱۳۹۰

شاخص تمیز سؤالات			شاخص دشواری سؤالات			نام گروه آموزشی
P-Value	شاخص تمیز سؤالات	P-Value	شاخص دشواری سؤالات	P-Value	شاخص دشواری سؤالات	
	مناسب **	آسان	دشوار	مناسب *	آسان	دارای دستیاری
	دشوار					گروه جراحی
	٪.۷۳/۵	٪.۱۰/۳	٪.۲۰/۶	٪.۶۷/۳	٪.۱۲/۱	گروه داخلی
	٪.۸/۴	٪.۷۰/۳	٪.۲۱/۳	٪.۶۰/۲	٪.۲۰/۵	گروه اطفال
۰/۰۰۱			۰/۰۰۱			گروه بیهوشی
	٪.۴۸/۹	٪.۵۱/۱	٪.۱/۳	٪.۳۴/۴	٪.۶۴/۳	گروه زنان
	٪.۴۲/۵	٪.۴۲/۵	٪.۱۶	٪.۳۲	٪.۵۲	جمع کل
	٪.۳۷/۸	٪.۴۹/۳	٪.۱۱/۵	٪.۲۹/۵	٪.۵۹	
	٪.۱۲/۹	٪.۳۴/۹	٪.۱۳/۷	٪.۴۴/۷	٪.۴۱/۶	
	٪.۱۰/۵					

* منظور سؤالاتی است که در محدوده (۰/۳ - ۰/۵) و به عنوان سؤالات نه سخت و نه آسان ارزیابی شدند.

** منظور سؤالاتی است که در محدوده (۰/۲۵ - ۰/۳۵) بوده و به عنوان سؤالات مناسب برای تشخیص دانشجویان قوی از ضعیف می‌باشد.

بحث

استفاده از گزینه منفی در ساقه سؤال منفی در آزمون‌های برگزار شده بیش از سایر اشکالات ساختاری بود که با نتیجه انجام شده در دانشگاه علوم پزشکی اراک تطبیق نداشت (۲۶،۲۵،۲۳) بدیهی است وجود گزینه منفی در ساقه سؤال منفی باعث سردرگمی دانشجو در انتخاب گزینه صحیح خواهد شد که در نهایت اعتبار ابزار ارزیابی دانشجو را زیر سؤال خواهد برد. به نظر می‌رسد واحد تایپ سؤالات، توسط مدیر محترم دستیاری، توجیه نشده و کنترل و نظارت‌های لازم توسط طراحان سؤال، قبل از تکثیر سؤالات، با دقت انجام نشده است. این پژوهش با تحقیق شکورنیا و همکاران همخوانی نداشت. چون بیشترین مشکلات مورد بررسی در پژوهش وی مربوط به وجود ابهام و عدم وضوح در هدف و موضوع در ساقه سؤالات چندگزینه‌ای بود که در ٪.۱۵/۹۴ سؤالات، این ابهام وجود داشت (۲۱) در مطالعه تارانت نیز بیشترین اشکالات ساختاری سؤالات مربوط به عدم تبیین کامل هدف در ساقه سؤال ارزیابی گردید (۲۴) در حالی که بیشترین اشکالات ساختاری در آزمون ارتقاء دستیاران در

در این پژوهش، سؤالات آزمون ارتقاء دستیاران این دانشگاه از جنبه‌های مختلفی مورد ارزیابی قرار گرفت که نتایج مبین این واقعیت می‌باشد که اعضای هیئت علمی دانشگاه علوم پزشکی اراک در طراحی سؤالاتی که دارای حداقل اشکالات ساختاری باشد، موفق بودند که با نتایج پژوهشی که توسط شکورنیا و همکاران در یک مطالعه مشابه که در دانشگاه جندی شاپور اهواز انجام گردید نیز همخوانی داشت. به طوری که در تحقیق شکورنیا نیز فقط ٪.۳۴/۸ سؤالات چند گزینه‌ای آزمون ارتقاء دستیاران دارای اشکالات ساختاری بود (۲۱) همچنین این پژوهش با تحقیقات انجام شده توسط واعظی، حق شناس و عرب منطبق بود (۲۳،۲۲،۱۳) اما با تحقیق انجام شده توسط تارانت و همکاران (Tarrant & et al) در دانشکده پزشکی دانشگاه هنگ‌کنگ که در سال ۲۰۰۶ میلادی انجام پذیرفت، همخوانی نداشت. چون تحقیق وی نشان داد که حدود ٪.۵۰ سؤالات طراحی شده دارای اشکالات ساختاری بودند (۲۴). تحقیقات انجام شده در دانشگاه‌های علوم پزشکی همدان، اهواز و یزد نیز نشان داد که

را از نظر شاخص تمیز (تشخیص) نامناسب ارزیابی نمود که با سوالات گروه‌های اطفال، بیهوشی و زنان در این تحقیق نیز همخوانی داشت (۳۲) همچنین با تحقیق انجام شده توسط درخشان و همکاران همخوانی نداشت. چون در تحقیق وی فقط ۱۶/۳٪ سوالات دانشکده پزشکی قزوین دارای ضریب دشواری و تمیز مناسب بودند (۳۳). به نظر می‌رسد علت طرح سوالات استاندارد در گروه‌های جراحی و داخلی این دانشگاه مربوط به بالا بودن میزان تجارب علمی و عملی اعضای هیئت علمی این گروه‌ها و نیز عضویت برخی از آنان در کمیته طراحی سوالات در سطح وزارتخانه می‌باشد. گرچه تجهیز واحد ارزیابی آزمون به دستگاه‌های الکترونیکی و تحلیل شاخص‌های دشواری و تمیز سوالات می‌تواند این مشکل را برای طراحی آزمون سال‌های آتی، برطرف نماید.

همچنین این تحقیق نشان داد در کل گروه‌های آموزشی فقط ۴۲٪ سوالات جزء اهداف اصلی آموزش بوده و ۳۲/۸٪ و ۳۲/۴٪ سوالات به ترتیب در سطح کاربرد و فهم از حیثه شناختی بلوم قرار داشته است که علت این امر عدم دارا بودن بانک سوالات استاندارد و نیز عدم طراحی سوالات به شکل محرمانه، هم‌زمان با تدوین طرح درس توسط اساتید بالینی دارای دستیار بوده است. در تحقیق انجام شده توسط دل آرام نیز طراحی سوالات آزمون توسط اساتید، قبل از برگزاری آزمون به عنوان یک نکته منفی در استانداردسازی آزمون ذکر شده است (۳۴) در تحقیق حق شناس و همکاران (۱۳۸۶) نیز بیش از سه چهارم سوالات در تاکسونومی I (یادآوری) طراحی شده بود و در مطالعه رسولی نژاد و همکاران نیز اکثریت سوالات آزمون ارتقاء دستیاران در کاشان هم در سطح تاکسونومی I قرار گرفته بود که با نتایج این تحقیق، همخوانی داشت (۳۵،۲۲).

از محدودیت‌های این پژوهش، ضعف در آگاهی برخی از اساتید گروه‌های آموزشی و نیز برخی مسئولین، از فرآیند ارزیابی آزمون ارتقاء و به تبع آن، تمایل اندک برای دریافت بازخورد از نتایج این ارزیابی بود. همچنین وجود نقص فنی در

دانشگاه علوم پزشکی اراک مربوط به عدم همسنگ بودن گزینه‌ها از نظر طول ساختاری و نیز سبک نگارش بود که در ۶/۷٪ سوالات این اشکال مشهود بود که به نظر می‌رسد پایین بودن سطح مهارت طراحان سؤال و نیز عدم نظارت کافی بر طراحی سوالات و عدم ارائه بازخورد به گروه‌ها، دلیل وجود اشکالات ساختاری بوده است. در هر حال، وجود اشکالات ساختاری به هر میزان، روایی سازه را در آزمون مختل نموده و در این صورت کارکردهای آموزشی آزمون را تحت تأثیر قرار داده و این مسأله با ایجاد زمینه برای کسب نمره با روش‌هایی غیر از تسلط علمی (حدس زدن، گمراه شدن دانشجویان و ...) نگرش دانشجویان را نسبت به آزمون با اشکال مواجه می‌سازد. با تلاش در زمینه ارتقاء جنبه‌های اصلی روایی محتوا و ساختار آزمون‌های چند گزینه‌ای، این آزمون‌ها به عنوان مطلوب‌ترین شیوه‌های ارزیابی دانش دانشجویان تبدیل می‌شوند. ماویس و همکاران (Mavis & et al) از دانشگاه میشیگان نشان دادند تنها در صورت استفاده از سوالات چندگزینه‌ای به عنوان یک روش رایج، اعتبار قابل توجهی برای آزمون‌ها ایجاد می‌کند (۲۷). مطالعه جواد و همکاران در بیم نشان داد که شایع‌ترین اشکالات در سوالات، مربوط به درج مطالب تکراری در گزینه‌ها و وجود لغات منفی در ساقه سوالات است که با این تحقیق همخوانی نداشت (۲۸) جوزیفوویز (Jozefowicz) نیز وجود اشکالات ساختاری در آزمون‌های اعضای هیئت علمی دانشگاه‌ها را عدم اختصاص فرصت کافی در طراحی سوالات ذکر می‌نماید. او همچنین بیان می‌کند که اساتید علیرغم صرف زمان زیاد برای آماده سازی مطالب درسی، زمان اندکی را برای طراحی سوالات استاندارد صرف می‌نمایند (۲۹). در تحقیق مک کوبری و همکاران (McCaubri & et al) در دانشگاه بریستول و نیز تحقیق هاموند و همکاران (Hammond & et al) از دانشگاه هامپتون (Hampton)، وجود اشکالات ساختاری در درصد قابل توجهی از سوالات نشان داده شد (۳۱،۳۰). پرویزی در ارزشیابی آزمون دکترای پرستاری، سوالات را از نظر ضریب دشواری مناسب ارزیابی نمود که با آزمون گروه‌های جراحی و داخلی تطبیق داشت ولی پرویزی سوالات

سپاس و قدردانی

بدینوسیله از زحمات سرکار خانم لیلا رحمتی (کارشناس ارشد برنامه ریزی آموزشی)، جناب آقای سید امین حسینی (کارشناس امور دستیاران)، جناب آقای دکتر محمود امینی (مدیر گروه جراحی)، جناب آقای دکتر حسین سرمیدیان (ریاست دانشکده پزشکی)، جناب آقای دکتر علی ارجمند شبستری (معاون بالینی دانشکده پزشکی)، جناب آقای دکتر محمد خلیلی (مدیر گروه بیهوشی)، سرکار خانم دکتر مهتری جمیلیان (مدیر گروه زنان)، سرکار خانم دکتر منیژه کهبازی (مدیر گروه اطفال)، سرکار خانم دکتر افسانه طلایی (مدیر گروه داخلی) که اینجانب را در انجام این پژوهش یاری دادند، نهایت تشکر و سپاسگزاری را می‌نمایم.

سیستم‌های الکترونیکی تحلیل آزمون در سال ۱۳۹۰ در دانشکده، منجر شد تا زمان انجام این تحقیق طولانی شود.

نتیجه‌گیری

نتایج این پژوهش بر لزوم تشکیل و تقویت تیم‌های ارزیابی آزمون در دانشگاه‌های علوم پزشکی و نیز ارتقاء سیستم‌های الکترونیکی تحلیل آزمون و نیز انجام مشاوره با کارشناسان تعلیم و تربیت و روان سنجی آزمون را تأکید می‌نماید و برگزاری کارگاه‌های آموزشی اثربخش و نیز کنترل سؤالات قبل از برگزاری آزمون‌ها و بالاخره اطلاع رسانی پیرامون نتایج آنالیز سؤالات و اعمال سیاست‌های تشویقی در گروه‌های آموزشی توانمند را با هدف طراحی استانداردترین سؤالات، به عنوان اثربخش‌ترین استراتژی‌ها ذکر می‌نماید.

References

- 1- Bahraini-Tosi MH, Modirazizi MJ, Kaveh Tabatabaee M, Ebraheemzadeh S. Bahraini Tosi V, Bahraini Tosi K. Clinical evaluation and medical school students in 2003 . Iran J Med Education 2002; 2:19. [Persian]
- 2- Mohamadee Zeidee A, Heidarnia AR. Khalaj M. Comparison of questions designed to test health school before and after accreditation workshop in 2008. Proceedings of the Ninth National Conference on Medical Education-Yazd 2008. [Persian]
- 3- Joint Center for Medical Education in Great Britain (JCEM). Assessment guide for practitioners and physician assistants. Translated by Asefzadeh S. Journal of Qazvin University of Medical Sciences 2001:58. [Persian]
- 4- Saberian M, Hajee Aghajane S. Curriculum Process in Medical Sciences, Tehran, Salemi Press. 2007:234. [Persian]
- 5- Abasi A, Kamkar A, Mahamed F. Assessment of multiple choice questions at the end of the nursing and midwifery schools' program in 2006-2007, Proceedings of the Ninth National Conference on Medical Education Yazd 2008:183. [Persian]
- 6- Yoshida M ,Tanaka M, Mizuno K, Ishii A, Nozaki K, Urakawa A. Factors influencing the academic motivation of individual college students. Int J Neurosci.2008;118(10):1400-11.
- 7- Bloom BS. Hasting HT. Madaus GF. Handbook of Formative and Summative Evaluation, New York, McGraw-Hill, 1971.

- 8- Gronlund N.E. Linn RL. Measurement & Evaluation in teaching; 7th Ed. New York. MacMillan, 1997.
- 9- Flahatee M, Mohammadzadeh Z, Entezare MH. Improving the quality and validity of test administration in the testing sessions. Iran J Med Education 2006; 5(1):5-6. [Persian]
- 10- Epstein RM. Assessment in Medical Education. J Med Edu, 2007; 356(4): 387-96.
- 11- Garakyaraghi M, Nazem M. Ebrahimi A. Avizhgan M. Mazaheri M. Ismaili A. The effect of an intervention designed to improve academic achievements questions in 2008-2009. Proceedings of the Ninth National Conference on Medical Education-Yazd 2008. [Persian]
- 12- Khoshrang H, Taheri M, Asadi A. Assessment quality of resident enhances questions in educational departments in Gilan University of Medical Sciences. Proceedings of the 4th National Festival of Shaheed Motahari-Mashhad 2012. [Persian]
- 13- Vaezi AA, Morovvati MA, Heydari MR. Procurement rules in multiple questions chooses at the end of the semester faculties. Proceedings of the 9th National Conference on Medical Education-Yazd 2008. [Persian]
- 14- Bazargan A. Educational Evaluation: Concepts, models and Operational processes. Tehran, Publications Samt; 2002:9-17. [Persian]
- 15- Kaveh Tabatabaee M, Bahraini-Tosi MH, Derakhshan A. Clinical evaluation and medical school students in 2003. J Mashhad University of Medical Sciences 2003; 45(76):89-95. [Persian]
- 16- James SC. Designing test for evaluation student achievement, Landon, Longman, 1990: 27.
- 17- Rezaee HR, Shoghee Shafagharya F, Doolati Y, Mohaghegh MA, Malekzadeh R. Assessment techniques in medicine. Tehran, Ameh Press; 1998. [Persian]
- 18- Alami A, Derakhshan F, Safdari F. Principles and design and evaluation of a multiple choose questions. Journal of Qazvin University of Medical Sciences; 2012:1. [Persian]
- 19- Amin MM, Shayan Sh, Hashemi H, Poursafa P, Ebrahimi A. Analysis of multiple choice questions based on classical test theory. Iran J Med Education 2007; 5(10):719-25. [Persian]
- 20- Saif AA. Measurement, assessment and evaluation of training. Tehran, Dooran Press; 2006. [Persian]
- 21- Shakurnia A, Khosravi Borojeni A, Mozaffari AR, Elhampour H. An evaluation of exam questions designed by faculty members, emphasizing on the multiple choice questions structure. Journal of Ahvaz University of Medical Sciences; 2009:6(2):129-138. [Persian]
- 22- Haghshenas MR, Vaheed Shahi K, Mahmoodi M, Shahbaznezhad L, Nezhad Neekoo P, Emadee A. Survey of multiple choice questions in Mazandaran University of Medical Sciences in 2007-2008; Journal of Mazandaran University of Medical Sciences 2008:5(2):119-126. [Persian]
- 23- Arab M. Survey of multiple choice questions defects in the Hamadan University of Medical Sciences in 2000. Proceedings of the 4th National

- Conference on Medical Education. Tehran 2003. [Persian]
- 24- Tarrant M , Knierim A, Hayes SK, Ware J. The frequency of item Writing flaws in multiple choice questions used in high stakes nursing assessment. *Nur Educ Today* 2006,26(8), 662-71.
- 25- Latifi M, Shakurnia A. Compare opinions of professors and experts about multiple choice questions in Ahvaz University of Medical Sciences. Proceeding of the 8th National Conference on Medical Education-Kerman 2007. [Persian]
- 26- Mohammadi Zeidi A, Heidarnia AR, Khalaj M. Comparison of questions designed to test the Department of Health before and after standardization test workshop. Proceedings of the Nnth National Conference on Medical Education Yazd 2007. [Persian]
- 27- Mavis DE, Cole BL, Hoppe RB. Survey of student assessment in US medical schools: The balance of breadth and fidelity. *Teach Learn Med* 2001; 13(2):74-9.
- 28- Javadi M, Abaszadeh A, Borhani F, Abdoli R. Analytical study of an end of term multiple choice questions in Bam Nursing School. Proceedings of the 8th National Conference on Medical Education-Kerman 2007. [Persian]
- 29- Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH, The Quality of in -house medical school examination. *Acad Med* 2002, 77(2):156-61.
- 30- McCouribe P. Improving the fairness of multiple choice questions: A literature review. *Med Teach* 2004:26(8):709-12.
- 31- Hammond EJ, Mcindo AK, Sansome AJ, Spargo PM. Multiple choice questions examination: Adapting evidence based approach to exam technique. *Anesthesia* 1998: 53(11): 1105-8.
- 32- Parvizi S, Peirovi H. Evaluation of 'Ph.D of Nursing' Entrance Exam. *Iran J Med Education* 2002; 8:43. [Persian]
- 33- Derkhan F, Jahan Hashemi H. Evaluation of tests at the end semester in medical schools. Proceedings of the 10th National Conference on Medical Education-Shiraz 2009:308-390 [Persian]
- 34- Delaram M, Foorozandeh N. Survey of student evaluation by faculties in Sharkord University of Medical Sciences. *Strides in Development of Medical Education* 2010; 7(1):51-56. [Persian]
- 35- Rasolinezhad A, Vakili Z, Fakharian A, Mosayebi F, Moniri R, Sayah M, Momen Heravi M, Soleimani AR, Tfasi Z, Saoori Kashani M. Comparison of taxonomic tests assistants of Kashan University of Medical Sciences in 2006. Proceedings of the 10th National Conference on Medical Education-Kerman 2007. [Persian]

Assessment of the resident's promotion exam: One step to validity of competency measurement in Arak University of Medical Sciences

z. Anbari (PhD)^{*1}, R. Jadidi (PhD)²

1, 2- Department of EDC, Arak University of Medical Sciences, Arak, Iran

Received: 24 Jul 2012

Accepted: 17 Nov 2012

Abstract

Introduction: Designing a tool for measuring of residents' competency with attention to their main role in education and practice of university. This study aims to assess the residents' promotion tests of clinical departments at Arak University of Medical Sciences.

Methods: This cross-sectional study that was undertaken in 2010 at Arak University of Medical Sciences. Seven hundred and fifty multiple choice questions related to resident promotion tests in surgery, internal medicine, pediatrics, gynecology and anesthesiology was compared. Questionnaire of each department contained 150 questions.

These questions were evaluated in the following domains: structure, Blum taxonomy, discrimination and difficulty index of questions and compliance to the core curriculums. Data gathering tool were: Millmen standard check list for evaluating questions' structure and check list for evaluating Blum taxonomy and core curriculum and OMR system for evaluating discrimination and difficulty index. The validity and reliability of tools was confirmed and data were analyzed using by ANOVA and X2 tests.

Results: Results showed gynecology department had structural problem (4.5 ± 4.2) compared with other departments. Internal medicine department had the highest Blum domain (40% application and 47% comprehension), surgery department had the highest learning aims (90.7%) and was assessed as the most suitable questions from difficulty index (67.3%) and discrimination index (73.5%). There was significant difference between structural problem, core curriculum and rate of standard questions in various clinical departments ($P=0.001$).

Conclusion: This study confirmed the necessity of test assessment in universities, to form effective educational workshops, control of questions before exams and incentives for clinical departments to design standard questions. Development of electronic question analysis system is recommended.

KeyWords: Comparison, exam, resident's promotion, university of medical sciences

*Corresponding author's email: dr.anbari@arakmu.ac.ir

This paper should be cited as:

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