

ارزیابی سؤالات چندگزینه‌ای پایان‌ترم دانشجویان پرستاری دانشگاه علوم پزشکی ارومیه

رحیم بقایی^۱، ژاله نادری^۲، شمس‌الدین شمس^۳، آرام فیضی^۴، داود رسولی^{۵*}

تاریخ دریافت ۱۳۹۴/۱۱/۱۶ تاریخ پذیرش ۱۳۹۵/۰۱/۲۰

چکیده

پیش‌زمینه و هدف: یکی از روش‌های متداول در ارزیابی دانشجویان استفاده از آزمون‌های چندگزینه‌ای و به‌طور رایج آزمون‌های چهارگزینه‌ای هست. هدف از مطالعه حاضر بررسی میزان رعایت شاخص‌های امتحانات چندگزینه‌ای پایان‌ترم دانشجویان پرستاری دانشگاه علوم پزشکی ارومیه در سال ۱۳۹۳ می‌باشد. **مواد و روش کار:** پژوهش حاضر مطالعه‌ای توصیفی می‌باشد که به‌صورت مقطعی در سال ۱۳۹۳-۹۴ بر روی ۷۰۵ نمونه از سؤالات چندگزینه‌ای پایان‌ترم اعضای هیئت‌علمی گروه پرستاری دانشکده پرستاری و مامایی ارومیه انجام گرفته است. کلیه سؤالات از نظر تاکسونومی بلوم در سه سطح تاکسونومی یک (دانش و فهم)، دو (کاربرد و آنالیز) و سه (ترکیب و ارزشیابی) موردبررسی قرار گرفتند. برای محاسبه شاخص‌های سؤالات از فرمول‌های ضریب دشواری و ضریب تمیز استفاده گردید. ساختار آزمون نیز با استفاده از چک‌لیست بررسی سؤالات چندگزینه‌ای میلن تحلیل گردید.

یافته‌ها: در مطالعه حاضر ۷۰۵ سؤال موردبررسی قرار گرفت که گروه داخلی - جراحی بیشترین تعداد سؤالات را به خود اختصاص می‌داد. اکثر سؤالات (۸۴/۶ درصد) دارای تاکسونومی یک بودند. اکثر سؤالات موردبررسی دارای ضریب دشواری خیلی سخت بودند (۳۳۲ سؤال) و در میان دروس بررسی‌شده داخلی - جراحی ۳ دارای بیشترین دشواری بود (۶۱/۴۲ درصد) و درس پرستاری مادران و نوزادان (۲) دارای کم‌ترین سختی بود (۱۰ درصد). در رابطه با ضریب تمیز سؤالات اکثر سؤالات دارای ضریب تمیز متوسط بودند (۲۹/۳۶ درصد) و در میان دروس بررسی‌شده پرستاری بیماری‌های روان (۱) دارای بهترین ضریب تمیز بود. اکثر سؤالات (۸۴/۱ درصد) دارای ساختار مناسب بودند.

نتیجه‌گیری: نتایج مطالعه حاضر نشان داد که سؤالات موردبررسی از لحاظ تاکسونومی بلوم، ضریب دشواری و ضریب تمیز در شرایط مطلوب نبوده و نیازمند برنامه‌ریزی مسئولان مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی ارومیه برای آموزش و توانمندسازی اساتید در زمینه‌های ذکرشده می‌باشد.

کلیدواژه‌ها: آزمون‌های چندگزینه‌ای، تاکسونومی، چک‌لیست میلن، آموزش پرستاری، ضریب دشواری، ضریب تمیز

مجله دانشکده پرستاری و مامایی ارومیه، دوره چهاردهم، شماره چهارم، پی‌درپی ۸۱، تیر ۱۳۹۵، ص ۲۹۹-۲۹۱

آدرس مکاتبه: دانشکده آموزش علوم پزشکی شهید بهشتی، دانشگاه علوم پزشکی شهید بهشتی، تلفن: ۰۲۱-۲۶۲۱۰۰۹۲

Email: dr.rasouli24@gmail.com

مقدمه

این جهت امتحانات درسی می‌باشند که به‌صورت دوره‌ای در فواصل مشخصی از هر نیمسال تحصیلی توسط متصدیان و متولیان امر آموزشی دانشگاه برگزار می‌شوند. باید در نظر داشت که ارزشیابی یکی از مهم‌ترین مراحل آموزشی است که بخش عمده‌ای از وظائف مدرسین در امر آموزش را در بر دارد (۱، ۲). برگزاری آزمون به‌منظور سنجش میزان صلاحیت و توانمندی افراد دارای پیشینه تاریخی می‌باشد آن چنانکه مقامات چینی در ۲۲۰۰ سال قبل از میلاد از

یکی از مهارت‌های مهم دانشگاه آموزش اثربخش دانشجویان است لذا در همین راستا برنامه‌هایی به‌منظور تربیت و پرورش بهینه دانشجویان طرح‌ریزی می‌گردد. برون داد این برنامه‌های آموزشی توسط شاخص‌هایی به‌منظور سنجش میزان نیل به اهداف آموزشی توسط ابزارهایی که برنامه نویسان و سیاست‌گذاران آموزشی هر کشور طراحی می‌کنند، ارزیابی می‌گردد. ازجمله ابزارهای رایج و شایع در

^۱ استادیار، مرکز تحقیقات و ایمنی بیمار، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی ارومیه

^۲ مرکز تحقیقات و ایمنی بیمار، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی ارومیه

^۳ مربی دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی ارومیه

^۴ استادیار، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی ارومیه

^۵ دانشجوی دکتری تخصصی آموزش پزشکی، دانشکده آموزش علوم پزشکی شهید بهشتی، دانشگاه علوم پزشکی شهید بهشتی (نویسنده مسئول)

شاخص‌های فوق‌الذکر به بیراهه می‌روند و این امر موجب عدم نیل به اهداف آموزشی موردنظر می‌شود. (۱۱) همچنین اهمیت طراحی آزمون‌های کارا مبین میزان پیشرفت تحصیلی فراگیران به‌قدری است که متولیان نظام آموزشی کشورها همواره بخشی از برنامه‌ها و تلاش‌های خود را جهت ارزیابی خود آزمون‌ها به کار می‌برند (۱۲). کلام آخر اینکه بدون شک نوع و کیفیت امتحان مستقیماً بر روش تدریس و به‌طور کلی بر اعتبار معلم تأثیر می‌گذارد. لذا در تهیه و اجرای آزمون‌ها، شایسته است که دقت و صداقت لازم به عمل آید تا آزمون از ویژگی‌های مطلوب برخوردار بوده و بتواند به‌عنوان یک ابزار پایا و معتبر جهت ارزیابی فراگیران که آینده‌سازان زمین‌های مختلف کشور می‌باشند، بکار گرفته شود (۱۳). با توجه به مغفول ماندن این حلقه‌ی اساسی در زنجیره‌ی آموزش پزشکی کشور، در این پژوهش، پژوهشگران برآنند که پژوهش حاضر را در جهت ارزیابی امتحانات چندگزینه‌ای دانشجویان دانشگاه علوم پزشکی ارومیه انجام بدهند.

مواد و روش کار

پژوهش حاضر مطالعه‌ای توصیفی می‌باشد که به‌صورت مقطعی در سال ۱۳۹۳-۹۴ بر روی ۷۰۵ نمونه از سؤالات چندگزینه‌ای پایان‌ترم اعضای هیئت‌علمی گروه پرستاری دانشکده پرستاری و مامایی ارومیه انجام گرفته است. پس از برگزاری آزمون پاسخ‌نامه دانشجویان برای تجزیه و تحلیل از اساتید مربوطه گرفته شد. کلیه سؤالات از نظر تاکسونومی بلوم در سه سطح تاکسونومی یک (دانش و فهم)، دو (کاربرد و آنالیز) و سه (ترکیب و ارزشیابی) موردبررسی قرار گرفتند. برای محاسبه شاخص‌های سؤالات از فرمول‌های ضریب دشواری^۱ و ضریب تمیز^۲ استفاده گردید. ضریب دشواری عبارت است از درصدی از افراد که به یک سؤال پاسخ درست داده‌اند و با P نشان داده می‌شود و مقدار آن بین صفر تا یک می‌باشد. در مطالعه حاضر ضریب دشواری ۰ تا ۰/۳ خیلی سخت، ۰/۳ تا ۰/۷ مناسب و ۰/۷ تا ۱ خیلی آسان در نظر گرفته شد. ضریب تمیز به توانایی سؤال در جدا کردن افراد قوی از ضعیف اشاره دارد و با D نشان داده می‌شود و مقدار آن بین ۱- تا ۱+ می‌باشد. هرچه این ضریب بزرگ‌تر باشد بهتر است. در پژوهش حاضر ضریب تمیز بزرگ‌تر از ۰/۴ خوب، ۰/۱۲ تا ۰/۳۹ متوسط، ۰ تا ۰/۱۱ بد و کم‌تر از صفر (منفی) در نظر گرفته شد. ساختار آزمون نیز با استفاده از چک‌لیست بررسی سؤالات چندگزینه‌ای میلمن تحلیل گردید. امتیاز کسب‌شده در این سیاهه بین ۱۴ (حداقل نمره) تا ۲۸ (حداکثر نمره) محاسبه شد که امتیاز ۱۴ به معنی عدم وجود ساختار و امتیاز ۲۸ به معنی ساختار

آزمون‌های کتبی برای گزینش افراد لایق جهت احراز تصدی امور مملکتی استفاده می‌کردند (۳) همچنین در ایران باستان در عهد شاپور پادشاه سامانی در دانشگاه جندی‌شاپور هم جلساتی برای آزمون دانشجویان پزشکی برپا می‌کردند و اعطای گواهینامه پزشکی موقوف به موفقیت دانشجویان در این آزمون بود (۴). بعلاوه از مطالعه آثار و نوشته‌های بزرگان کشورمان چنین استنباط می‌شود که آن‌ها هم برای ارزشیابی اندوخته‌های شاگردان و طلاب حوزه‌ی درسی خود از آزمون به‌صورت شفاهی در محضر استاد یا کتبی استفاده می‌کردند (۵). امروزه بیش از ۲۰ میلیون کودک، نوجوان و جوان در مقاطع مختلف به تحصیل اشتغال دارند که بیش از دو میلیون نفر آن‌ها در دانشگاه‌ها و مؤسسات آموزش عالی کشور تحصیل می‌کنند. برای تصمیم‌گیری در مورد میزان پیشرفت و ارتقای آنان به‌طور مرسوم اساتید از انواع مختلف آزمون‌ها استفاده می‌کنند. جهت اطمینان از اینکه این فراگیران به یادگیری مطلوب دست پیدا کرده‌اند و ارزشیابی آن‌ها از عینیت، اعتبار و ثبات لازم برخوردار می‌باشد، استفاده از آزمون‌هایی است که توسط اعضای هیئت‌علمی دانشگاه‌ها مؤسسات آموزش عالی ساخته می‌شود، تنها وسیله‌ای است که می‌توان به آن اطمینان کرد (۶). ارزشیابی به سبب ماهیت و پیچیدگی‌های خاص خود از گسترده‌ترین و جنجالی‌ترین موضوعات آموزشی است و در صورتی که به نحو صحیحی اجرا شود می‌توان بر اساس نتایج آن نقاط قوت و ضعف آموزشی را مشخص نمود (۷). آزمون‌های مختلفی جهت سنجش میزان یادگیری فراگیران استفاده می‌شوند، در این میان سؤالات چندگزینه‌ای که اولین بار در سال ۱۹۱۴ در دانشگاه کانزاس توسط فردریک ابداع گردید، اکنون رایج‌ترین نوع آزمون‌ها در آموزش پزشکی می‌باشند (۸). نکته مهم و قابل‌ملاحظه در تدوین این سؤالات توجه به ویژگی‌های اساسی آن‌ها جهت ارزیابی است که عبارت‌اند از: اعتبار قابل‌قبول، روایی مناسب و علمی بودن. به همین منظور پارامترهایی شامل درجه دشواری سؤال، قدرت تمیز سؤال در تشخیص دانشجوی قوی و ضعیف، درصد پاسخ به هر یک از گزینه‌های انحرافی و صحیح همراه با اعتبار و خطای استاندارد استفاده می‌گردد (۹). معمولاً طراحان سؤال مهارت چندانی در طرح سؤال ندارند و این پیش‌داوری برای دانشجویان به وجود می‌آید که حتماً در هر امتحان چندگزینه‌ای سؤالاتی مبهم و دوپهلوی وجود دارد، اما این مسئله مربوط به طراحی سؤال است نه خود سؤال (۱۰). اهمیت طراحی و برگزاری آزمون‌های چندگزینه‌ای مطلوب در زمینه انتخاب سطح دانش (تاکسونومی) و رعایت قواعد ساختاری به‌قدری است که گهگاه مدرسان به دلیل برآورده نکردن مناسب هرکدام از

² Discrimination index

¹ Difficulty index

سؤالات). درس داخلی - جراحی ۲ دارای بیشترین تعداد سؤالات با ضریب دشواری متوسط بود (۳۲/۸۵ درصد سؤالات) و همچنین درس پرستاری مادران نوزادان (۲) آسان‌ترین سؤالات را ارائه کرده بود (۷۳/۳۳ درصد سؤالات). همچنین متوسط ضریب دشواری سؤالات ۰/۴۲ درصد بود. (جدول ۲).

در رابطه با ضریب تمیز سؤالات اکثر سؤالات دارای ضریب تمیز متوسط بودند (۲۹/۳۶ درصد) و در میان دروس بررسی‌شده پرستاری بیماری‌های روان (۱) دارای بهترین ضریب تمیز بود (۸۷/۵ درصد سؤالات) و درس پرستاری داخلی جراحی (۳) دارای بیشترین ضریب تمیز منفی بود (۲۴/۲۸ درصد سؤالات). درس داخلی - جراحی (۱) دارای بیشترین تعداد سؤالات با ضریب تمیز متوسط بود (۴۵ درصد سؤالات). متوسط ضریب تمیز سؤالات ۰/۱۸ درصد بود. (جدول ۳).

برای تعیین روایی سازه آزمون‌های چندگزینه‌ای مورد مطالعه از اصول میلمن استفاده گردید. ۱/۳ درصد از سؤالات دارای ساختار ضعیف، ۱۴/۶ درصد ساختار متوسط و ۸۴/۱ درصد ساختار خوب داشتند. بهترین ساختار آزمون مربوط به گروه پرستاری داخلی جراحی و پرستاری بیماری‌های روان (۱) بود. جدول شماره چهار، درصد اشکال ساختاری آزمون‌های چهارگزینه‌ای را به‌طور کلی در آزمون‌های مورد بررسی نشان می‌دهد.

کاملاً صحیح سؤالات بود. نمرات کسب‌شده به سه گروه دارای ساختار ضعیف (نمره ۱۸-۱۴)، متوسط (نمره ۲۳-۱۹) و خوب (نمره ۲۸-۲۴) طبقه‌بندی شدند. کلیه ملاحظات اخلاقی از جمله محرمانه ماندن اطلاعات و رضایت آگاهانه رعایت گردید. پس از ورود داده‌ها به نرم‌افزار SPSS مشخصات روان‌سنجی آزمون‌ها اعم از ضریب دشواری، ضریب تمیز آزمون بررسی گردید.

یافته‌ها

در مطالعه حاضر ۷۰۵ سؤال مورد بررسی قرار گرفت که گروه داخلی - جراحی ۳۹۰ (۵۵/۴ درصد)، کودکان ۷۰ (۹/۹ درصد)، بهداشت جامعه ۸۵ (۱۲/۰۹ درصد)، روان پرستاری ۷۰ (۹/۹ درصد)، مادران نوزادان ۶۰ (۸/۵ درصد) و مدیریت ۳۰ (۴/۲ درصد) از سؤالات را به خود اختصاص می‌دادند (جدول شماره ۱). از نظر تاکسونومی ۸۴/۶ درصد از سؤالات دارای تاکسونومی یک (دانش و فهم)، ۱۳/۵ درصد تاکسونومی دو (کاربرد و آنالیز) و ۱/۹ درصد دارای تاکسونومی سه (ترکیب و ارزشیابی) بودند. اکثر سؤالات مورد بررسی دارای ضریب دشواری خیلی سخت بودند (۳۳۲ سؤال) و در میان دروس بررسی‌شده داخلی - جراحی ۳ دارای بیشترین دشواری بود (۶۱/۴۲ درصد سؤالات) و درس پرستاری مادران و نوزادان (۲) دارای کم‌ترین سختی بود (۱۰ درصد

جدول (۱): عناوین دروس مورد بررسی و تعداد سؤالات و درصد آن‌ها

درصد سؤال	تعداد سؤال	درس
۸/۵	۶۰	داخلی - جراحی (۱)
۹/۹	۷۰	داخلی - جراحی (۲)
۹/۹	۷۰	داخلی - جراحی (۳)
۹/۳	۶۵	داخلی - جراحی (۴)
۹/۳	۶۵	پرستاری مراقبت‌های ویژه
۴/۳	۳۰	پرستاری کودکان (۱)
۵/۶	۴۰	پرستاری کودکان (۲)
۴/۳	۳۰	پرستاری بهداشت جامعه (۱)
۴/۳	۳۰	پرستاری بهداشت جامعه (۲)
۳/۵	۲۵	پرستاری بهداشت جامعه (۳)
۴/۳	۳۰	پرستاری بحران
۴/۳	۳۰	بررسی وضعیت سلامت
۴/۳	۳۰	مدیریت در پرستاری
۲/۷	۲۰	پرستاری مادران نوزادان (۱)
۵/۶	۴۰	پرستاری مادران نوزادان (۲)
۵/۶	۴۰	پرستاری بیماری‌های روان (۱)
۴/۳	۳۰	پرستاری بیماری‌های روان (۲)

جدول (۲): مقایسه میانگین و توزیع فراوانی دشواری سؤالات آزمون‌ها

متوسط دشواری	$P > 0.07$ (آسان)		$0.03 < P < 0.07$ (مناسب)		$P < 0.03$ (سخت)		درس
	درصد	تعداد	درصد	تعداد	درصد	تعداد	
۰/۳۷	۲۱/۶۷	۱۳	۲۳/۳۳	۱۴	۵۵	۳۳	داخلی - جراحی (۱)
۰/۳۸	۲۱/۴۳	۱۵	۳۲/۸۵	۲۳	۴۵/۷۲	۳۲	داخلی - جراحی (۲)
۰/۲۵	۱۸/۵۸	۱۳	۲۰	۱۴	۶۱/۴۲	۴۳	داخلی - جراحی (۳)
۰/۳۶	۲۲/۰۸	۱۵	۱۸/۴۶	۱۲	۵۸/۴۶	۳۸	داخلی - جراحی (۴)
۰/۴۲	۲۷/۶۹	۱۸	۲۷/۶۹	۱۸	۴۴/۶۲	۲۹	پرستاری مراقبت‌های ویژه
۰/۳۸	۲۶/۶۷	۸	۱۳/۳۳	۴	۶۰	۱۸	پرستاری کودکان (۱)
۰/۴۷	۳۰	۱۲	۳۵	۱۴	۳۵	۱۴	پرستاری کودکان (۲)
۰/۴۵	۲۰	۶	۳۶/۶۶	۱۱	۴۳/۳۴	۱۳	پرستاری بهداشت جامعه (۱)
۰/۵۰	۴۲/۳۳	۱۳	۳۳/۳۳	۱۰	۲۳/۳۴	۷	پرستاری بهداشت جامعه (۲)
۰/۴۸	۱۲	۳	۴۰	۱۰	۴۸	۱۲	پرستاری بهداشت جامعه (۳)
۰/۴۵	۱۳/۳۳	۴	۳۳/۳۳	۱۰	۵۳/۳۴	۱۶	پرستاری بحران
۰/۵۸	۷۲/۳۳	۲۲	۱۶/۶۷	۵	۱۰	۳	بررسی وضعیت سلامت
۰/۴۸	۳۰	۹	۴۳/۳۳	۱۳	۲۶/۶۷	۸	مدیریت در پرستاری
۰/۴۶	۲۵	۵	۵۰	۱۰	۲۵	۵	پرستاری مادران نوزادان (۱)
۰/۶۰	۸۰	۳۲	۱۰	۴	۱۰	۴	پرستاری مادران نوزادان (۲)
۰/۲۶	۰/۰	۰	۱۲/۵	۵	۸۷/۵	۳۵	پرستاری بیماری‌های روان (۱)
۰/۳۱	۶/۶۷	۲	۲۰	۶	۷۲/۳۳	۲۲	پرستاری بیماری‌های روان (۲)
۰/۴۲	۲۶/۹۵	۱۹۰	۲۵/۹۵	۱۸۳	۴۷/۱	۳۳۲	کل

جدول (۳): مقایسه میانگین و توزیع فراوانی درجه تمییز سؤالات آزمون‌ها

متوسط تمییز	$D < 0.4$ (منفی)		$0.4 < D < 0.7$ (ضعیف)		$0.7 < D < 0.9$ (متوسط)		$D > 0.9$ (خوب)		درس
	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	
۰/۲۲	۱/۶۸	۱	۲۶/۶۶	۱۶	۴۵	۲۷	۲۶/۶۶	۱۶	داخلی - جراحی (۱)
۰/۲۴	۲/۸۵	۲	۲۸/۵۷	۲۰	۳۸/۵۷	۲۷	۳۰	۲۱	داخلی - جراحی (۲)
۰/۱۸	۲۴/۲۸	۱۷	۲۲/۸۶	۱۶	۲۲/۸۶	۱۶	۳۰	۲۱	داخلی - جراحی (۳)
۰/۲۸	۱۲/۳۰	۸	۱۵/۳۹	۱۰	۲۱/۵۴	۱۴	۵۰/۷۷	۳۳	داخلی - جراحی (۴)
۰/۱۶	۱۳/۸۵	۹	۳۳/۸۵	۲۲	۳۲/۳۰	۲۱	۲۰	۱۳	پرستاری مراقبت‌های ویژه
۰/۲۱	۱۰	۳	۲۶/۶۶	۸	۳۳/۳۴	۱۰	۳۰	۹	پرستاری کودکان (۱)
۰/۱۷	۲۵	۱۰	۴۵	۱۸	۱۵	۶	۱۵	۶	پرستاری کودکان (۲)
۰/۲۶	۱۳/۳۳	۴	۳۰	۹	۴۰	۱۲	۱۶/۶۷	۵	پرستاری بهداشت جامعه (۱)
۰/۱۴	۱۶/۶۷	۵	۴۰	۱۲	۳۰	۹	۱۳/۳۳	۴	پرستاری بهداشت جامعه (۲)
۰/۲۰	۲۸	۷	۲۴	۶	۲۴	۶	۲۴	۶	پرستاری بهداشت جامعه (۳)
۰/۱۶	۳۰	۹	۳/۳۳	۱	۲۶/۶۷	۸	۴۰	۱۲	پرستاری بحران
۰/۱۳	۱۶/۶۷	۵	۵۰	۱۵	۲۰	۶	۱۳/۳۳	۴	بررسی وضعیت سلامت
۰/۱۱	۱۳/۳۳	۴	۴۰	۱۲	۳۳/۳۴	۱۰	۱۳/۳۳	۴	مدیریت در پرستاری
۰/۱۰	۱۰	۲	۵۵	۱۱	۳۰	۶	۵	۱	پرستاری مادران نوزادان (۱)
۰/۱۳	۵	۲	۴۷/۵	۱۹	۴۲/۵	۱۷	۵	۲	پرستاری مادران نوزادان (۲)
۰/۳۱	۰/۰	۰	۵	۲	۵	۲	۹۰	۳۶	پرستاری بیماری‌های روان (۱)
۰/۲۱	۶/۶۷	۲	۲۶/۶۷	۸	۳۳/۳۳	۱۰	۳۳/۳۳	۱۰	پرستاری بیماری‌های روان (۲)
۰/۱۸	۱۲/۷۷	۹۰	۲۹/۰۸	۲۰۵	۲۹/۳۶	۲۰۷	۲۸/۷۹	۲۰۳	کل

جدول (۴): توزیع فراوانی اشکالات ساختاری آزمون‌های چهارگزینه‌ای

ردیف	موضوع	بله (درصد)	خیر (درصد)
۱	آیا بخش اعظم اطلاعات در ساقه سؤال گنجانده شده است؟	۶۸/۵	۳۱/۵
۲	آیا سؤال یک هدف اختصاصی یادگیری را مورد ارزیابی قرار می‌دهد؟	۹۱/۱	۸/۹
۳	آیا لغات استفاده‌شده در ساقه یا گزینه‌ها، شفاف و مستقیم بیان‌شده‌اند؟	۹۰/۶	۹/۴
۴	آیا از کاربرد گزینه منفی برای ساقه منفی خودداری شده؟	۹۵/۴	۴/۶
۵	آیا از کاربرد گزینه‌هایی نظیر "همه موارد"، "هیچ‌کدام" و گزینه‌های ترکیبی خودداری شده است؟	۹۱/۹	۸/۱
۶	آیا از کاربرد گزینه‌های متضاد یکدیگر، خودداری شده است؟	۹۸	۲
۷	آیا از لغات مثبت در ساقه سؤال استفاده شده است یا در صورت منفی بودن ساقه سؤال، لغات منفی مشخص شده‌اند؟	۹۶/۷	۳/۳
۸	آیا این سؤال مستقل از سؤالات دیگر می‌باشد؟	۷۹/۴	۲۰/۶
۹	آیا گزینه‌ها از نظر طول، ساختار لغوی و سبک نگارش همسنگ هستند؟	۶۲/۹	۳۷/۱
۱۰	آیا تا حد امکان از کاربرد عبارت‌های تکراری در گزینه‌ها خودداری شده است؟	۶۵/۵	۳۴/۵
۱۱	آیا کلمات به‌کاررفته در ساقه و یا گزینه‌ها از نظر املايي صحیح است؟	۹۰	۱۰
۱۲	آیا گزینه‌ها به‌طور عمودی لیست شده است؟	۴۰/۶	۵۹/۴
۱۳	آیا به‌طور مشخص یک گزینه صحیح می‌باشد؟ خودداری از طرح سؤالات Controversy	۸۶/۱	۲۳/۹
۱۴	آیا ساقه سؤال و گزینه‌ها مثبت‌اند؟	۸۵/۳	۱۴/۷
۱۵	آیا از دادن سرنخ‌هایی که نشان‌دهنده گزینه صحیح می‌باشد، خودداری شده است؟	۸۹/۷	۱۰/۳

بحث و نتیجه‌گیری

می‌دهد، از آنجایی که اکثر سؤالات بررسی‌شده در مطالعه حاضر و مطالعات مرور شده دارای تاکسونومی سطح یک بودند، بنابراین مراکز توسعه آموزش باید تدابیر و روش‌های لازم برای توانمندسازی اساتید در طراحی سؤالات با تاکسونومی بالاتر را بکار گیرند. مطالعه حاضر نشان داد که اکثر سؤالات موردبررسی دارای ضریب دشواری بالا بودند. در مطالعه کلهری و همکاران اکثر سؤالات دستیاری دارای ضریب دشواری متوسط بودند (۱۶). در مطالعه وفامهر و همکاران بیشتر سؤالات دارای ضریب دشواری آسان بودند (۱۹). اشرف پور و همکاران در مطالعه خود در دانشگاه علوم پزشکی بابل بیان کردند که تنها ۳۸ درصد از سؤالات دارای ضریب دشواری مناسب هستند. تحلیل آزمون‌های ارتقاء رشته‌های پزشکی نشان داد که تنها ۵۴ درصد سؤالات دارای ضریب دشواری قابل‌قبول می‌باشند و مابقی سؤالات یا بسیار آسان طرح شده‌اند یا بسیار سخت (۲۰). تشنیزی و همکاران نیز در مطالعه خود به این نتیجه رسیدند که اکثر سؤالات دارای ضریب دشواری مناسب می‌باشند و پس‌از آن بیشتر سؤالات ضریب دشواری کم داشته و آسان طراحی شده‌اند (۲۱). میترا و همکاران در پژوهش خود نشان دادند که ۸۰ درصد سؤالات دارای ضریب دشواری کم بوده و برای دانشجویان آسان بودند (۲۲). جوادی مهر و همکاران در مطالعه خود بر روی سؤالات نیمسال دوم سال تحصیلی ۲۰۱۰-۲۰۱۱ نشان دادند که ۴۵/۶ درصد سؤالات دارای ضریب دشواری کم بوده و ۴۰ درصد دارای

در مطالعه حاضر تاکسونومی اکثر سؤالات در سطح I بود. در مطالعه حق‌شناس و همکاران نیز اکثر سؤالات موردبررسی دارای تاکسونومی یک بوده و دانش، فهم دانشجویان را می‌سنجیدند (۱۴). رسولی نژاد و همکاران نیز در مطالعه خود بر روی سؤالات آزمون ارتقاء دستیاری اکثر سؤالات را در تاکسونومی سطح یک (یادآوری) گزارش کردند (۱۵). کلهری و همکاران نیز در مطالعه خود به این نتیجه رسیدند که اکثر سؤالات دستیاری دارای تاکسونومی سطح یک (یادآوری) می‌باشند (۱۶). سیاح و همکاران با مقایسه آزمون‌های ارتقاء و دانشنامه رشته‌های جراحی عمومی، زنان و مامایی، کودکان و داخلی نشان دادند که اکثر آزمون‌های ارتقا رشته داخلی و دانشنامه رشته جراحی عمومی دارای تاکسونومی سطح دو و سه می‌باشند. همچنین بیان کردند که سطوح تاکسونومی سؤالات نسبت به سال‌های گذشته خود یک روند توسعه‌ای را داشتند (۱۷). بولتر و همکاران در مطالعه خود نشان دادند که اکثر سؤالات موردبررسی دارای تاکسونومی سطوح دو و سه می‌باشند (۱۸). نتایج متناقض در سطوح تاکسونومی سؤالات بررسی‌شده می‌تواند به نحوی بیانگر توسعه توانایی طراحان سؤال در طراحی سؤالات بهتر در سال‌های اخیر باشد. پیشرفت روزافزون روش‌های ارزیابی دانشجویان و توانایی استفاده از روش‌های مختلف برای ارزیابی دانشجویان لزوم استفاده از سؤالات چندگزینه‌ای با تاکسونومی بالا را به‌خوبی نشان

چیدمان عمودی گزینه‌ها، نبود بخش اعظم اطلاعات در ساقه سؤال، عبارات تکراری در گزینه‌ها، طول و ساختار لغوی، سبک نگارش و همسنگ نبودن گزینه‌ها بود (۲۶). بر اساس نتایج مطالعه سیاح ملی و همکاران ۷۲/۷-۵۵ درصد سؤالات در آزمون‌های ارتقاء و ۷۲/۷-۷۰/۲ درصد در آزمون‌های دانشنامه بدون اشکالات ساختاری بودند. در مطالعه آن‌ها رایج‌ترین اشکالات ساختاری مربوط به آیتم‌های نداشتن بخش اعظم اطلاعات در ساقه‌ی سؤال و عدم امکان پاسخ دادن به سؤالات بدون مراجعه به گزینه‌ها بود (۱۷). حق‌شناس و همکاران نیز بیان کردند که ۴۶ درصد از سؤالات موردبررسی بدون اشکالات ساختاری بوده و بقیه اشکالات ساختاری جزئی داشتند (۱۴). در مطالعه شکوری نیا و همکاران نیز اکثر سؤالات ساختار خوب داشتند (۲۷). نتایج مطالعه جوادی مهر و همکاران نیز همخوان با یافته‌های پژوهش حاضر بود (۲۳). واعظی و همکاران نشان دادند که تنها ۲۴ درصد از سؤالات اعضای هیئت‌علمی دانشکده‌های پزشکی، پرستاری و بهداشت دارای ساختار خوب بودند. در این میان دانشکده پرستاری و مامایی نسبت به بقیه دانشکده‌ها بیشترین نمره را کسب کرده بود (۲۸). Tarranet و همکاران نشان دادند ۴۶/۲ درصد از سؤالات چندگزینه‌ای به‌کاررفته در ارزشیابی دانشجویان دارای اشکالات ساختاری بوده است (۲۹). رشته‌های پزشکی و پیراپزشکی به دلیل حساس بودن نیازمند آموزش و ارزیابی دقیق می‌باشد، چراکه فارغ‌التحصیلان رشته‌های مذکور مستقیماً با سلامتی افراد جامعه درگیر می‌باشند. نتایج مطالعه حاضر نشان داد که سؤالات موردبررسی از لحاظ تاکسونومی بلوم، ضریب دشواری و ضریب تمیز در شرایط مطلوب نبوده و نیازمند برنامه‌ریزی مسئولان مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی ارومیه برای آموزش و توانمندسازی اساتید در زمینه‌های ذکرشده می‌باشد.

تشکر و قدردانی

پژوهش حاضر طرح مصوب دانشگاه علوم پزشکی ارومیه با شماره ۴۰۸-۳۳-۰۲-۹۰ می‌باشد. بدین‌وسیله پژوهشگران مراتب تشکر و قدردانی خود را از مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی ارومیه به دلیل حمایت مالی، دادن مجوز انجام مطالعه، از اعضای محترم هیئت‌علمی دانشکده پرستاری و مامایی ارومیه به دلیل شرکت در این مطالعه را اعلام می‌دارند.

ضریب دشواری مناسب بودند (۲۳). Si-Mui Sim و همکاران نشان دادند که ۷۵ درصد سؤالات موردبررسی دارای ضریب دشواری کم بودند (۲۴). تناقضات موجود در مطالعات بررسی‌شده می‌تواند به دلیل ماهیت‌های متفاوت سؤالات موردبررسی باشد ولی آن چیزی که در اکثر آن‌ها به‌طور مشترک مشاهده می‌شود نیاز به آموزش و ارتقاء اعضای هیئت‌علمی برای طراحی سؤالات با ضریب دشواری مناسب بخصوص در رشته‌های پزشکی و پیراپزشکی می‌باشد.

مطالعه حاضر نشان داد که اکثر سؤالات موردبررسی دارای ضریب تمیز متوسط بودند. در مطالعه کلهر و همکاران نیز ضریب تمیز اکثر سؤالات موردبررسی متوسط بود (۱۶). در مطالعه تشنیزی و همکاران نیز ۵۷/۷ درصد سؤالات دارای ضریب تمیز مناسب بودند (۲۰) در مطالعه اشرف پور و همکاران متوسط درجه تمیز سؤالات ۰/۱۴ به‌دست آمد که نشان می‌دهد سؤالات از قدرت تمیز مناسبی برخوردار نیستند (۲۱). در مطالعه میترا و همکاران ۶۷ درصد سؤالات دارای ضریب تمیز بالای ۰/۲ و در حد قابل‌قبول بودند (۲۲). در شبان و همکاران قدرت تمیز ۴۰/۶ درصد سؤالات حتی بعد از مداخله آموزشی پایین‌تر از ۰/۲ بود (۲۵). در مطالعه ثناگو و همکاران ضریب تمیز هر ۱۲ آزمون موردبررسی کم بود و بیشترین قدرت تمیز ۰/۳۲ بود (۲۶). در مطالعه جوادی مهر و همکاران در دانشگاه علوم پزشکی زاهدان اکثر سؤالات دارای ضریب تمیز بالاتر از ۰/۳ بودند. همچنین در مطالعه آن‌ها ۱۸/۲ درصد سؤالات ضریب تمیز منفی داشتند (۲۳). نتایج متفاوت در مطالعات بررسی‌شده می‌تواند به دلیل ماهیت‌های متفاوت مطالعات در سؤالات بررسی‌شده باشد. در رابطه با ضریب تمیز نیز همانند ضریب دشواری نیاز به آموزش و ارتقاء اعضای هیئت‌علمی برای طراحی سؤالات با ضریب تمیز بالاتر به‌خوبی احساس می‌شود.

در مطالعه حاضر اکثر سؤالات بر اساس سیاهه بررسی ساختار سؤالات چندگزینه‌ای میلن دارای ساختار خوب بودند که نشان از توانمند بودن اعضای هیئت‌علمی دانشکده پرستاری و مامایی ارومیه در طراحی سؤالات می‌باشد. بیشترین اشکالات ساختاری سؤالات بررسی‌شده در " آیا گزینه‌ها به‌طور عمودی لیست شده است؟" و " آیا گزینه‌ها از نظر طول، ساختار لغوی و سبک نگارش همسنگ هستند؟" بود. در مطالعه ثناگو و همکاران نیز اکثر سؤالات موردبررسی بر اساس سیاهه بررسی ساختار سؤالات چندگزینه‌ای میلن ساختار خوب داشتند. در مطالعه آن‌ها هیچ سؤالی با ساختار ضعیف شناسایی نشد. بیشترین اشکالات ساختاری مربوط به

References:

1. Seif A. Educational measurement, assessment and evaluation. 4th ed. Tehran: Doran; 2004.P. 426-8. (Persian)
2. Arab M. Survey of defects of multiple choice questions examinations of Medical School in second half of 1999. Journal of Tehran medical faculty, Special Fourth National Conference on Medical Education, 2000. (Persian)
3. Ebel, R.L. and Frisbie, D.A. (5th ed). (1991) Essentials of Educational Measurement. Englewood Cliffs, NJ: Prentice Hall.
4. Kyamanesh A. multiple-choice Test and its impact on student learning, educational content and teaching methods. Juornal of education, number 31. (Persian)
5. Isa S. Iran culture history. Tehran university publication, number 424. (Persian)
6. Sepasi h, Atari yousef A. the study of psychometric characteristics of shahid chamran university faculty members' final test scores. JOURNAL OF EDUCATION AND PSYCHOLOGY WINTER 2006, Volume 12, Number 4; Page(s) 1 - 20. (Persian)
7. Falahati M, Mohamadzadeh Z, Entezari M. Improvement of quality and validity of student assemssment: executive aspects. Iranian Journal of Medical Education. 2005; 5 (1): 5-6. (Persian)
8. Kaveh M. Analytical assessment of multiple – choice test at medical school. Medical Journal of Mashhad Univerity of Medical Sciences 2002; 76 (45): 89-95. (Persian)
9. Saburi M, Shayan S, Salehi A, Honarpisheh H. Review Psychometric Parameters of the 29th Residency Test (1380) According to the Classic Test Theory (CTT). Iranian Journal of Medical Education. 2002; 2: 49-50. (Persian)
10. Zolfaghari B, Asadollahi GH. Academic achievement tests in medical sciences. Isfahan: Isfahan University of Medical Sciences; 2000, 35-52. (Persian)
11. Schotanus C. Student assessment and examination rules. Medical Teacher 1999; 21(3): 318-21.
12. Najar S, Abedi P. The survey of knowledge and practice of faculty members about validity and reliability of exams. Iranian Journal of Medical Education. 2002; 2: 42-42. (Persian)
13. Shabani H. Educational skills. Samt publivation, 1993: 315-394. (Persian)
14. Haghshenas M, Vahid Shahi K, mahmudi M, Shahbazenejad L, Parvinne jad 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 Eduction 2008; 5(2): 120-127. (Persian)
15. Rasolinejad SA, Vakihi Z, Fakharion E, Mosayebi Z, Moniri R. Comparative survey of taxonomies of residents. Promotion examination, kashan Medical University 2006. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 68. (Persian)
16. Pourmirza Kalhori R, Rezaie M, Shojee Moghadam AR, Sepahi V, Memar Eftekhari L. Correlation of quality and quantity index of multiple choice questions exams of residency promotion in Kermanshah University of medical sciences, 2013. J Clin Res Paramed Sci 2015; 4(1): 71-78. (Persian)
17. Sayyahmelli M, Barzegar M, Bilan N, Aslanabadi S, khoshbaten M, Gasemzadeh A, Karegar Maher MH. Comparing Multiple-Choice Questions Quality Parameters of Pediatric, General Surgery, Internal Medicine and Genecology and Obstetrics Residency Tests in Preboard and Promotion Examinations of Tabriz University of Medical Sciences and National Board Examination in 2010 and 2011. Journal of Medical Education Development, 2015; 8(18): 43-54. (Persian)

18. Butler AC, Korpicke JD, Roediger HL. Correcting metacognitive error: feedback increases retention of low- confidence correct response. *J Exp psychol learn Mem Cogn*. 2008 Jul; 34(4): 418-2.
19. Vafamehr V, Dadgostarnia M. Reviewing the results of qualitative and quantitative analysis of MCQs in Introduction to clinical medicine course. *IJME* 2010; 10(5): 1146-1152. [Persian]
20. Ashrafpour M, Beheshti Z, Molookzadeh F. Quality of final examination in students of Babol medical university, 1999-2000. *Journal of Babol University of Medical Sciences*. 2003; 0: 42- 7. (Persian)
21. Teshnizi SH, Zare S, Solati M. Quality analysis of multiple choice questions (MCQs) examinations of noncontinuous undergraduate medical records. *medical journal of hormozgan*; 2010, Issue 3: page 177-183
22. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The Levels of Difficulty and Discrimination Indices in Type a Multiple Choice Questions Of Preclinical Semester 1, Multidisciplinary Summative Tests. *IeJSME* 2009; 3 (1): 2-7.
23. Mishmast Nehy GA, Javadimeh M. Analysis of Multiple Choice Questions of the Second Semester Examinations Held in Zahedan College of Nursing and Midwifery in 2010-2011. *Life Sci J* 2013; 10(3): 1045-1051.
24. Sim SM, Isaiah Rasiah R. Relationship between Item Difficulty and Discrimination Indices in True/FalseType Multiple Choice Questions of a Para-clinical Multidisciplinary Paper. *Ann Acad Med Singapore* 2006; 35: 67-71.
25. Shaban M, Ramazani Badr F. Effect of test item analysis on summative exams quantity of test designing. *HAYAT*; 2007; 13(1); 5-15. (Persian)
26. Sanagoo A, Jouybari L, Ghanbari Gorji M. Quantitative and qualitative analysis of achievement tests in Golestan University of Medical Sciences. *J Research in Medical Education*. 2010; 2(2): 24-32. (Persian)
27. Shakurnia AH, Mozaffari AR, Khosravi Brougeni. Survey on structural of MCQs of residency exam in AJUMS. *Judishapour Scientific Medical Journal*. 2009; 8(4): 491-502. (Persian)
28. Vaezi AK, Morovati MA, Heidari MR. Evaluation multiple choice questions in final semester exams. *Journal of Medical Education Development*; issue 2: page 31-38. (Persian)
29. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessment. *Medical Education*. 2008; 42(2): 198-206.

EVALUATION OF THE NURSING STUDENTS FINAL EXAM MULTIPLE-CHOICE QUESTIONS IN URMIA UNIVERSITY OF MEDICAL SCIENCES

Rahim Baghaei¹, Jaleh Naderi², Shamsedin Shams³, Aram Feizi⁴, Davood Rasouli^{5}*

Received: 6 Feb, 2016; Accepted: 9 Apr, 2016

Abstract

Background & Aims: One of the common methods of student assessment tests is using multiple choice and four-choice tests. The aim of this study was to observe the parameters of nursing student's summative multiple-choice exams in the Urmia University of Medical Sciences in 2014.

Materials & Methods: This descriptive cross-sectional study was carried out on a sample of 705 final exam multiple-choice questions at the Faculty of Nursing and Midwifery in 2014-15. All questions evaluated in three levels of Bloom's Taxonomy; one (knowledge and understanding), two (Application and Analysis), and three (synthesis and evaluation). Formulas were used to calculate the indices of questions difficulty coefficient and discrimination coefficient. The structure of tests was analyzed by using the Mailman multiple-choice checklist.

Results: Out of 705 questions which were examined in this study, the majority belonged to the Department of Medical-Surgical. Most questions (84/6 percent) had one taxonomy level. Considering difficulty index, most of the questions (332 items) were very difficult, and among the studied subjects, Medical-Surgical 3 was the most difficult (61/42 percent) and obstetric nursing (2) was the least difficult (10 percent). Related to the discrimination index, majority of questions had average discrimination coefficient (29/36 percent) and among the subjects studied mental illnesses nursing (1) had the best coefficient. Most questions (1.84 per cent) had appropriate structure.

Conclusion: This study showed that based on Bloom's Taxonomy, difficulty coefficient and discrimination coefficient wasn't in good condition in the Nursing Faculty of Urmia University of Medical Sciences, and it requires further planning by Medical Education Development Center..

Keywords: Multiple choice tests, Taxonomy, Mailman checklist, Nursing education, Difficulty index, Discrimination coefficient

Address: School of Medical Education, Shahid Beheshti University of Medical Sciences

Tel: (+98) 26210092

Email: dr.rasouli24@gmail.com

¹ PhD of Hospital management, Patient safety research center, School of Nursing and Midwifery, Urmia

² School of Nursing and Midwifery, Urmia University of Medical Sciences, Urmia

³ Instructor of Nursing and Midwifery, University of Medical Sciences, Urmia, Iran

⁴ Patient safety research center, School of Nursing and Midwifery, Urmia

⁵ PhD candidate of medical education, School of medical education, Shahid Beheshti University of Medical Sciences (Corresponding Author)