مقاله پژوهشی اصیل

مجلـــه‌ی توسـعه‌ی آمـوزش در علـوم پزشـــکی

دوره‌ی ۸، شماره ۲۰، زمستان ۹٤ ، صفحات ۲۵ تا ۳۲

ارزیابی کیفی و کمی سوالات آزمون ارتقای دستیاری سال های ۹۱ و ۹۲ دانشگاه علوم پزشکی قزوین

فریبا درخشان، دکتر سلیمان احمدی، دکتر عباس علامی

نویسنده‌ی مسوول: تهران، دانشگاه علوم پزشکی شهید بهشتی، دانشکده‌ی آموزش علوم پزشکی faribaderakhshan@gmail.com

دریافت: ۹۳/۱۱/۱۱ پذیرش: ۹٤/۵/۱۲

چکیده

زمینه و هدف: *طراحی دقیق سوالات آزمون ارتقا مبتنی با اصول استاندارد از اهمیت بالایی برخوردار است. هدف از این مطالعه ارزیابی کیفی و کمی سوالات آزمون ارتقای دستیاری سال های ۹۱ و ۹۲ دانشگاه علوم پزشکی قزوین و بررسی ارتباط بین سطوح تاکسونومی با شاخص های کمی سوالات آزمون ارتقای دستیاری بود.*

روش بررسی: *کلیه‌ی سوالات آزمون ارتقای دستیاری در ۷ رشته‌ی تخصصی پزشکی دانشگاه علوم پزشکی قزوین در سال‌های ۹۱ و ۹۲ مورد مطالعه قرار گرفت. تاکسونومی سوالات در سه سطح دانش، درک و کاربرد تعیین گردید. قدرت افتراق در چهار گروه خوب، متوسط، ضعیف و منفی بررسی شد. درجه‌ی دشواری در ۵ سطح ساده، متوسط، دشوار، خیلی دشوار و فوق العاده دشوار تقسیم بندی گردید. اطلاعات وارد نرم‌افزار SPSS v.19 شد.جهت بررسی ارتباط بین متغیرها از آزمون کای زوجی و ANOVA استفاده شد.*

یافته ها: *از مجموع ۲۱۰۰ سوال، ۳۸درصد تاکسونومی یادآوری، ۲٤/۱درصد در سطح درک و ۳۷/۹درصد در سطح کاربرد بود. قدرت افتراق ۲٤/۹ درصد سوالات منفی، ۳٤/۱ درصد ضعیف، ۲۲/۵ درصد متوسط و ۱۸/۵درصد خوب بود. ازنظر درجه دشواری، ۲۲/۱درصد سوالات ساده، ۲۲/٦درصد متوسط، ۲٤/۵درصد دشوار، ۱۷/۵درصد خیلی دشوار و ۱۳/۲درصد فوق‌العاده دشوار بود. ارتباط معناداری بین تاکسونومی و قدرت افتراق به‌دست آمد، سوالات با تاکسونومی یادآوری قدرت افتراق کمتر و در سوالات با تاکسونومی درک و کاربرد قدرت افتراق بالاتر بود($p<۰/۰۰۱$).*

نتیجه گیری: *افزایش سطح تاکسونومی سوالات می‌تواند باعث بهبود قدرت افتراق سوالات آزمون گردد. با توجه به این‌که سوالات طراحی‌شده با تاکسونومی دو و سه کمتر از حد استاندارد بود(سطح استاندارد در آزمون‌هایPost Graduate، ۸۰ درصد می‌باشد) پیشنهاد می‌گردد، اساتید در آزمون ارتقا، تعداد سوالات بیشتری با تاکسونومی بالاتر طرح نمایند.*

واژگان کلیدی: ***تاکسونومی، ضریب تمایز، ضریب دشواری، آزمون ارتقا، دستیار***

مقدمه

آزمون‌های با سوالات چندگزینه‌ای با برخورداری از عینیت (Objectivity) و پایایی (Reliability) بالا، از متداول‌ترین و رایج‌ترین نوع ارزشیابی آموزشی هستند و به دلیل سهولت نسبی اجرا به‌ویژه در مقیاس‌های بزرگ، واجد امتیازات

۱- کارشناس ارشد آموزش پزشکی، دانشکده آموزش علوم پزشکی، دانشگاه علوم پزشکی شهید بهشتی

۲- دکترای تخصصی آموزش پزشکی، استادیار دانشگاه علوم پزشکی شهید بهشتی

۳- متخصص عفونی، دانشیار دانشگاه علوم پزشکی قزوین

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

ویژه‌ای برای آزمون‌های مهم می‌باشند (۱). این نوع سوالات اولین بار در آزمون کتبی بوردهای تخصصی ایالات‌متحده مورداستفاده قرار گرفت (۲). در ایران نیز آزمون‌های چندگزینه‌ای متداول‌ترین آزمون کتبی در ارزشیابی دانشجویان گروه پزشکی است. درصورتی‌که سوالات چندگزینه‌ای به‌خوبی طراحی شوند می‌توانند سطوح دانش، فهم، ادراک، کاربرد و اطلاعات را نیز ارزشیابی کنند (۳). آنچه در سوالات چندگزینه‌ای مورد ارزیابی قرار می‌گیرد اهداف آموزشی حیطه‌ی شناختی است (٤). در حیطه‌ی شناختی از دانشجو می‌خواهیم موضوع آموزش داده‌شده را به یاد آورد، یا موضوع یا مشکل را تشخیص داده و یا تفسیر کند و یا راه‌حل مشکل یا مساله‌ای را بیان کند. با این نگاه سوالات چندگزینه‌ای را می‌توان به سه رده تاکسونومی یک (دانش)، دو (درک) و سه (کاربرد) تقسیم نمود. در آزمون‌های دوره های بالینی، پایه‌ی (stem) سوالات با سطح تاکسونومی دو و سه، معمولا معرفی یک بیمار است که شامل شکایت اصلی به همراه علایم و نشانه‌های مرتبط، اطلاعات آزمایشگاهی و نظیر آن می‌باشد. این موضوع که آیا سوالات چندگزینه‌ای برای سنجش دانش، در آزمونی در سطح دانشنامه تخصصی مناسب می باشند، همواره موردتوجه قرار داشته است. با وجود این که برخی نگرانی‌ها در مورد تناسب کاربرد این نوع سوالات در آزمون‌های پایانی تخصص‌های بالینی مطرح می شود با این حال اکثر مطالعات اخیر مانند مطالعه Downing و همکاران، Smith و همکاران و Ware و همکاران در سال ۲۰۰۸ ارزش کاربرد سوالات چندگزینه‌ای برای سنجش دانش در آزمون‌هایی در سطح دانشنامه و سطوح مشابه را اثبات نموده‌اند و ازاین‌رو زمینه‌ی مناسبی برای ادامه تلاش‌ها و انجام مطالعاتی مشابه فراهم گشته است (٦ و ٥).

سوالات یک آزمون چند گزینه ای علاوه بر اینکه قبل از برگزاری باید از نظر مشکلات نگارشی، ساختاری و اهداف مورد بازبینی قرار گیرد، پس از برگزاری آزمون نیز باید تمامی سوالات از نظر شاخص‌های پس‌آزمونی (Post Examination Survey) همچون میزان دشواری و قدرت افتراق بررسی شود (٤). Tyneو همکاران (۲۰۱۳) در تحقیقات خود نشان دادند ضریب دشواری و تاکسونومی سوالات دو فاکتور مهمی است که در روایی آزمون‌ها موثرند و از این‌رو باید در هر آزمونی توسط طراحان سوال موردتوجه قرار گیرند (۷) در بین این شاخص ها ضریب تمایز از اعتبار کافی برخوردار است (٤). این شاخص بیان‌گر آن است که سوال مزبور تا چه اندازه توانسته دانشجویان قوی (گروهی که بیشترین نمره را در مجموع کل آزمون کسب کرده اند) را از دانشجویان ضعیف (گروهی که کمترین نمره را در مجموع کل آزمون کسب کرده اند) متمایز نماید(٤). در بررسی سوالات چندگزینه ای توجه به تاکسونومی و قدرت افتراق سوالات از اهمیت خاصی برخوردار است. تحقیقات Kaufman و Van Der Vleuten نشان داد نگرانی‌های قابل‌توجهی در مورد برخی شاخص‌های کیفی سوالات چندگزینه‌ای شامل سطح دانش بکار رفته (تاکسونومی)، رعایت اصول ساختاری و هم‌چنین شاخص‌های پس‌آزمونی وجود دارد (۹ و ۸).

دبیرخانه شورای تخصصی برنامه گردآوری ۲۰ درصد برتر سوالات آزمون ارتقای دستیاران در رشته‌های مختلف در بانک سوال را در دستور کار خود قرار داده است و وزارت بهداشت نیز در استفاده بیشتر از سوالات با سطوح تاکسونومی بالاتر، توجه و اهتمام به طراحی و اجرای آزمون‌های چندگزینه‌ای استاندارد در آزمون‌های دستیاری تاکید دارد (۱۰). خوشرنگ و همکاران در مطالعه‌ای به بررسی کیفیت ۳٦۰۰ سوال آزمون ارتقای دستیاری در ۱۲ رشته‌ی تخصصی در دانشکده‌ی پزشکی گیلان طی سال‌های ۸۹ و ۹۰ پرداختند. نتایج این مطالعه نشان داد بیش از نیمی از سوالات در سطح تاکسونومی یک طراحی‌شده بود (۱۱). در دانشگاه علوم پزشکی مازندران در سال تحصیلی ۸٦-۸۵، ۷۷/٤ درصد سوالات در تاکسونومی ۱ و مابقی در تاکسونومی ۲ و ۳ بودند (۱۲). نتایج

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

سایر مطالعات در داخل کشور نیز حاکی از وجود نقص در طراحی سوالات چندگزینه‌ای آزمون‌های ارتقای دستیاری است (۱۳ و ۱٤ و ۱۵). از این رو مطالعه‌ی حاضر باهدف مقایسه‌ی نتایج تجزیه‌وتحلیل تاکسونومی و شاخص های پس آزمونی سوالات آزمون ارتقای دستیاری در دانشگاه علوم پزشکی قزوین در سال‌های ۹۲-۹۱ انجام پذیرفت.

روش بررسی

این پژوهش یک مطالعه‌ی توصیفی- تحلیلی بود. واحدهای مورد مطالعه، سوالات چهارگزینه‌ای مربوط به آزمون‌های کتبی دستیاری تخصصی پزشکی در ۷ رشته‌ی (اطفال، بیهوشی، جراحی، داخلی، روان‌پزشکی، زنان و زایمان و قلب) در دانشگاه علوم پزشکی قزوین بود. در این پژوهش کلیه‌ی سوالات آزمون‌های ارتقای دستیاری تخصصی پزشکی دانشگاه علوم پزشکی قزوین در سال‌های ۱۳۹۱ و ۱۳۹۲ مورد بررسی قرار گرفت. معیار خروج، آزمون‌های دستیاری که دریکی از این سال‌ها برگزار نشده است تعیین شد، لذا آزمون کتبی رشته تخصصی عفونی که در سال ۱۳۹۱ وجود نداشت از دور مطالعه خارج گردید. با توجه به این‌که هر آزمون ارتقا شامل ۱۵۰ سوال چهارگزینه‌ای بود، درمجموع ۲۱۰۰ سوال چهارگزینه‌ای یک پاسخ درست (Single Best Answer) در آزمون کتبی ارتقای دستیاران تخصصی مورد تجزیه‌وتحلیل قرار گرفت.

شاخص کیفی (تاکسونومی سوالات) توسط کارشناسان ارشد آموزش پزشکی مجرب موردبررسی قرار گرفت. روایی کار با توجه به سطوح شناختی در سه سطح ۱(یادآوری)،۲(تفسیر اطلاعات) و ۳(اقدام و حل مسئله) بررسی شد. برای تعیین پایایی، ابتدا جلسات هماهنگی بین دو نفر از کارشناسان آموزش پزشکی (که به اصول تاکسونومی سوالات آشنا بودند و یکی از آنان مدرس کارگاه مزبور در سطح دانشگاه بوده است) برگزار شد و توافق نظر اولیه در مورد نحوه‌ی بررسی سوالات ازلحاظ تاکسونومی به عمل آمد. همچنین دو مصحح که دوره کارشناسی ارشد آموزش پزشکی را گذرانده، در ابتدا و در طول مطالعه جلسات هماهنگی بین آنان برگزار شد و توافق نظر در مورد نحوه بررسی سوالات به عمل آمد. با بررسی حدود ۵۷ درصد سوالات (٦۰۰ سوال) توسط مصححین، میزان همخوانی نتایج بین ارزیابی کننده ها(Inter-Rater) ۸٦/۵ درصد به‌دست آمد. موارد مبهم در تشخیص سطوح تاکسونومی بین کارشناسان دوباره مطرح شد و توافق مشترک حاصل گردید.

اطلاعات مربوط به پاسخ‌نامه توسط مارک خوان به نرم‌افزار تجزیه‌وتحلیل یگانه منتقل‌شد و اطلاعات مربوط به شاخص کمی (ضریب تمایز) و قدرت افتراق و درجه دشواری سوالات با کمک نرم‌افزار محاسبه شد. ضریب تمایز با دامنه تغییرات ۱- تا ۱+ بود و قدرت افتراق در چهار گروه خوب، متوسط، ضعیف و منفی و درجه دشواری در ۵ سطح ساده، متوسط، دشوار، خیلی دشوار و فوق العاده دشوار تقسیم بندی گردید. ضریب تمایز بین ۱-تا ۰، ۰ تا ۰/۱، ۰/۱ تا ۰/۳، ۰/۳ تا ۱ به ترتیب دارای قدرت افتراق منفی، ضعیف، متوسط و خوب در نظر گرفته شد.

آنالیزهای اولیه در رابطه با ضریب تمایز سوالات محاسبه شد، سپس داده‌های هر دو دوره یعنی اطلاعات مربوط به ضریب تمایز، درجه دشواری و تاکسونومی سوالات، به تفکیک گروه‌های آموزشی وارد نرم‌افزار آماری SPSS.Ver.19 شد. مقایسه بین داده‌های کمی با استفاده از آزمون T زوجی و ANOVA مورد تجزیه‌وتحلیل قرار گرفت. خطای آلفا کمتر از ۰/۰۵ در نظر گرفته شد.

یافته‌ها

تعداد۱۰۵۰سوال مربوط به سال ۹۱ و ۱۰۵۰ سوال مربوط به سال ۹۲در مجموع ۲۱۰۰ سوال در ۷ رشته تخصصی اطفال، جراحی، داخلی، قلب، روانپزشکی، زنان، بیهوشی در

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

دانشگاه علوم پزشکی قزوین مورد تجزیه و تحلیل قرار گرفت. دانشگاه علوم پزشکی قزوین حدود ۳ سال است که در رشته‌های قلب و روان‌پزشکی به تربیت دستیار پرداخته است، لذا طراحی سوالات آزمون ارتقای دستیاری در گروه‌های آموزشی قلب و روان‌پزشکی تجربه‌ی کمتر از ۳ سال است ولی گروه‌های آموزشی اطفال، بیهوشی، زنان و جراحی و داخلی بیش از ۱۵ سال تجربه طراحی سوال در آزمون‌های ارتقای دستیاری را داشته‌اند.

در بررسی تاکسونومی سوالات، نتیجه تحلیل‌ها نشان داد ۳۸درصد سوالات درسطح تاکسونومی یادآوری (دانشی)، ۲٤/۱درصد در سطح درک و ۳۷/۹درصد در سطح کاربرد بوده است.

جدول ۱: فراوانی مطلق و نسبی تاکسونومی سوالات آزمون ارتقای دستیاری به تفکیک گروه‌های آموزشی سال‌های ۹۱ و ۹۲

گروه آموزشی	سال												p
	۹۱						۹۲						
	سطح تاکسونومی سووالات						سطح تاکسونومی سووالات						
	یک		دو		سه		یک		دو		سه		
اطفال	۵۲	۱۳/۹	۵۲	۱۸/۸	٤٦	۱۱/۵	٤۸	۱۱/۳	٤۷	۲۰/۳	۵۵	۱۳/۹	۰/۵٤۵
بیهوشی	٦۵	۱۷/٤	۵۷	۲۰/۷	۲۸	۷	۷۷	۱۸/۲	٤۵	۱۹/۵	۲۸	۷/۱	۰/۲۹۷
داخلی	۱٤	۳/۷	۵۵	۱۹/۹	۸۱	۲۰/۳	۲۲	۵/۲	۵٤	۲۳/٤	۷٤	۱۸/۷	۰/۳٤۹
جراحی	۲۵	٦/۷	۲۵	۹/۱	۱۰۰	۲۵	۱٤	۳/۳	۲۱	۹/۱	۱۱۵	۲۹/۱	۰/۱۰٦
قلب	۷۷	۲۰/٦	۳۸	۱۳/۸	۳۵	۸/۸	۹۰	۲۱/۲	۳٦	۱۵/٦	۲٤	٦/۱	۰/۲۱
روان	۱۰۸	۲۸/۹	۲۸	۱۰/۱	۱٤	۳/۵	۱۳٤	۳۱/٦	۱۲	۵/۲	٤	۱	۰/۰۰۱*
زنان	۳۳	۸/۸	۲۱	۷/٦	۹٦	۲٤	۳۹	۹/۲	۱٦	٦/۹	۹۵	۲٤/۱	۰/۵۵٤

P<۰/۰۵ * معنادار آماری

جدول شماره ۱ نشان می‌دهد ، اگرچه تفاوت معنی‌داری در تاکسونومی سوالات گروه روان‌پزشکی سال ۹۲ نسبت به سال ۹۱ وجود داشت (p=۰/۰۰۱)، با این حال در سال ۹۲ نیز بیشتر سوالات طراحی شده توسط گروه مزبور در سطح تاکسونومی یک بوده است. ازنظر تاکسونومی درون گروه‌های دیگر اختلاف معنادار آماری بین سوالات طراحی‌شده در سال‌های ۹۱ و ۹۲ یافت نشد. در بین گروه‌های آموزشی، گروه داخلی و جراحی به ترتیب با ۸۸ و ۸۷ درصد، دارای بهترین عملکرد و گروه روان‌پزشکی و قلب با ۱۹/۳ و ٤٤/۳ درصد سوالات با تاکسونومی دو و سه بیشترین تفاوت را با استاندارد تعیین شده در رده بندی تاکسونومی در بین گروه های آموزشی داشتند.

۲٤/۹ درصد سوالات دارای قدرت افتراق منفی، ۳٤/۱ درصد قدرت افتراق ضعیف، ۲۲/۵ درصد سوالات متوسط و ۱۸/۵

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

درصد سوالات قدرت افتراق خوب داشتند. میانگین ضریب تمایز ۰/۲۹ ± ۰/۰۹ در سال ۹۱ و ۰/۲۵± ۰/۰۸ در سال ۹۲ بود. میانگین ضریب تمایز ۰/۲۷± ۰/۰۸ به‌دست آمد. ازنظر درجه دشواری، ۲۲/۱درصد سوالات ساده، ۲۲/٦درصد متوسط، ۲٤/۵درصد دشوار، ۱۷/۵درصد خیلی دشوار و ۱۳/۲درصد فوق‌العاده دشوار و در کل ضریب دشواری با میانگین ۰/۲۹ ± ۰/٦۰ در سال ۹۱ و ۰/۲۷ ± ۰/۵۸ در سال ۹۲ به دست آمد.ضریب‌دشواری‌درسال‌های ۹۱ و ۹۲ اختلاف معنی‌داری نداشت. در مطالعه‌ی حاضر، رابطه‌ی معنـاداری بیـن ضریـب دشـواری و ضریب تمایز سوالات به‌دست آمد.

جدول ۲: فراوانی نسبی درجه دشواری سوالات آزمون ارتقا دستیاری به تفکیک گروه‌های آموزشی سال‌های ۹۱ و ۹۲

گروه آموزشی	سال											
	۹۱						۹۲					
	درجه دشواری						درجه دشواری					
	ساده	متوسط	دشوار	خیلی دشوار	فوق‌العاده دشوار	ساده	متوسط	دشوار	خیلی دشوار	فوق‌العاده دشوار	P
اطفال	۲٤/۷	۲٤/۷	۲٤/۷	۱۲/۷	۱۳/۳	۲٦/۷	۲۳/۳	۱٤/۰	۲۲/۰	۱٤/۰	۰/۰۷۹
بیهوشی	۱٤/۷	۱٦/۰	۳۵/۳	۲۰/۰	۱٤/۰	۱۸/۷	۲۷/۳	۲۰/۷	۲۱/۳	۱۲/۰	۰/۰۲٤*
داخلی	۱٦/۷	۲۱/۳	۳۰/۰	۲۲/۰	۱۰/۰	۸/۷	۳۲/۰	۱۷/۳	۲۷/۳	۱٤/۷	۰/۰۰۷*
جراحی	۰	۱۵/۳	۵۰/۷	۱۸/۷	۱۵/۳	۲۲/۷	۲۳/۳	۲٦/۰	۱۸/۷	۹/۳	۰/۰۰۰*
قلب	۳۰/۰	۳۰/۰	۱۸/۰	۱۳/۳	۸/۷	۳۱/۳	۳۲/۷	۱۵/۳	۱٦/۷	٤/۰	۰/٤۵۳
روان	۷۳/۳	۰	۱۸/۰	۰	۸/۷	۱۲/۷	۳۲/۰	۱٦/۷	۱۱/۳	۲۷/۳	۰/۰۰۰*
زنان	۸/۰	۱۸/۷	۳۲/۰	۲۵/۳	۱٦/۰	۲۱/۳	۲۰/۰	۲٤/۷	۱٦/۰	۱۸/۰	۰/۰۰۸*

P<۰/۰۵ * معنادار آماری

جدول ۲ نشان می دهد گروه روان پزشکی در سال ۹۱، ۷۳/۳ درصد و گروه قلب در سال ۹۲، ۳۰ درصد سوالاتی که طراحی کردند سوالات ساده بود که بالاترین درصد سوالات ساده را در بین گروه های آموزشی داشتند. ضریب دشواری به تفکیک گروه‌های آموزشی، در گروه‌های بیهوشی، داخلی، جراحی، روان‌پزشکی و زنان، در سال‌های ۹۱ و ۹۲ اختلاف آماری معنادار داشت (P<۰/۰۵). با استفاده از آزمون ANOVA ارتباط معناداری بین تاکسونومی سوالات و ضریب تمایز و درجه دشواری مشاهده شد (p <۰/۰۰۱).
همچنین بین سطوح تاکسونومی و قدرت افتراق سوالات با استفاده از آزمون کای زوجی ارتباط معنی‌دار بود (p<۰/۰۰۱) (جدول شماره ۳).

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

جدول ۳: ارتباط بین سطوح تاکسونومی و قدرت افتراق سوالات آزمون ارتقا دستیاری سال‌های ۹۱ و ۹۲

j قدرت افتراق	تاکسونومی			P
	۱	۲	۳	
منفی	۲۱۷	۱۲٤	۱۸۲	X^2= 72.53 Degrees of Freedom= 6 p-value <0.001
ضعیف	۳۲۵	۱۷۷	۲۱٤	
متوسط	۱۱۵	۱۱٤	۲٤۳	
خوب	۱٤۱	۹۲	۱٥٦	
جمع	۷۹۸	٥۰۷	۷۹٥	

بحث

گروه روان پزشکی و قلب در مقایسه با گروه های آموزشی دیگر کمترین تجربه طراحی سوال در آزمون ارتقا را داشته است، لذا گروه روان پزشکی در سال ۹۱ و گروه قلب در سال۹۲ بیشترین تعداد سوالات ساده را در مقایسه با گروه‌های آموزشی دیگر داشتند. در بررسی سطوح تاکسونومی سوالات، نیز گروه روان پزشکی در مقایسه با گروه های آموزشی دیگر بالاترین درصد سوالات در سطح تاکسونومی یک را داشته است. بررسی پژوهش حاضر و همچنین مطالعه‌ی پژوهش‌های انجام‌شده در داخل و خارج کشور نشان می‌دهد درصد سوالات طراحی‌شده در تاکسونومی دو و سه کم‌تر از سطح استاندارد است. درمجموع ٦۲ درصد سوالات مطالعه‌ی حاضر، در سطوح عمیق دانش (تاکسونومی دو و سه) طراحی‌شده است که کمتر از سطح استاندارد است (سطح استاندارد در آزمون‌هایpost graduate، ۸۰ درصد می‌باشد)(٤). نتایج پژوهش مسترز (Masters) و همکاران (سال ۲۰۰۱) در زمینه‌ی تاکسونومی نشان داد ٤۷/۳درصد سوالات موجود در بانک سوالات در حیطه‌ی دانش و آگاهی و ٦/٥ درصد در سطح تجزیه‌وتحلیل طراحی‌شده بود (۱٦). مطالعه هاکسهام (Huxham) و همکارانش در آمریکا در سال ۱۹۸۰ میلادی و تارانت (Tarrant) و همکارانش در سال ۲۰۰٦ از هنگ‌کنگ غلبه‌ی کاربرد تاکسونومی ۱ در آزمون‌های پیشرفت تحصیلی را نشان دادند (۱۸ و ۱۷). در تحقیقی که رسولی نژاد و همکاران انجام دادند اعلام داشتند که ٦٥/۸ درصد سوالات آزمون‌های دستیاران دانشگاه علوم پزشکی کاشان در سطح تاکسونومی یک بوده است (۱۹). مطالعات اخوان کرباسی بر روی ۱۱۰۳ سوال دستیاری گروه اطفال دانشگاه علوم پزشکی یزد در سال‌های ۱۳۸٤-۱۳۸۷ نشان داد نزدیک به نیمی از سوالات در سطح تاکسونومی یک طراحی‌شده بود (۲۰). بر اساس گزارش دبیرخانه شورای آموزش پزشکی و تخصصی وزارت بهداشت از مجموع ۱۱۲٥۰ سوال چهارگزینه‌ای مربوط به آزمون های کتبی دانشنامه ۲٥ رشته‌ی تخصصی طی سال‌های ۸۸-۸٦ میانگین تاکسونومی III و II، ۱۷/٦±٤٦/٥ به‌دست آمد (۲۱).

یکی از دلایل احتمالی غلبه سوالات در سطح محفوظات (تاکسونومی ۱) در آزمون‌ها، آسان تر بودن و اختصاص دادن زمان کمتر جهت طراحی سوالات با تاکسونومی یک در مقایسه با طراحی سوالاتی با تاکسونومی بالاتر است. سوالات با تاکسونومی پایین، موجب می‌شود دانشجویان به سمت یادگیری و کسب دانش سطحی و محفوظات روی‌آورند و از درک صحیح و تفسیر و حل مسئله غافل گردند(۲۲). در زمینه‌ی شاخص‌های کمی (ضریب دشواری و ضریب تمایز)،

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

یافته‌ها نشان داد بیش از یک سوم سوالات درجه‌ی دشواری ساده یا فوق‌العاده دشوار داشتند که در محدوده‌ی سوالات مناسب با ضریب دشواری ۰/۳ تا ۰/۷ قرار نگرفت، همچنین بیش از نیمی از سوالات (۵۹درصد) سوالات آزمون، قدرت افتراق منفی و ضعیف داشتند. در چهار گروه دستیاری دانشگاه علوم پزشکی کاشان، کمتر از ۵۰ درصد سوالات، ضریب تمایز مناسب داشتند (۲۳). نتایج مطالعه درخشان و همکاران نیز نشان داد تنها ۱۶/۳ درصد از مجموع ۱۵۸۶ سوال آزمون‌های دانشکده پزشکی قزوین در سال ۸۶ ازنظر ضریب دشواری و افتراق مناسب بودند (۲٤). در تحقیق شبستری و همکاران در دانشکده دندان‌پزشکی قزوین در سال‌های ۸۹ –۸۸ تنها یک‌سوم سوالات ضریب دشواری و افتراق قابل‌قبول داشتند (۲۵). نتایج مطالعه‌ای مشابه در دانشگاه علوم پزشکی یاسوج نیز موید ضعف قدرت افتراق سوالات طراحی شده در دانشگاه مذکور می باشد (۲٦). به نظر می‌رسد نتایج تحلیل آزمون‌ها، ارایه‌ی بازخورد به طراحان سوال و آموزش روش‌های مناسب طراحی سوال به ارتقای کیفیت سوالات آزمون کمک می کند(۲۸و۲۷،۱٤،۲۱).

در تحقیق حاضر، بین تاکسونومی و درجه دشواری سوالات ارتباط معنادار پیدا شد، بیشتر سوالات ساده تاکسونومی یک داشتند.تینگ(Tyng) و همکاران از مالزی در تحقیقی در سال ۲۰۱۳ بین پیچیدگی سوالات بر اساس طبقه‌بندی بلوم و دشواری سوالات ارتباط هرچند ضعیف یافتند(۷). در مطالعه‌ی حاضر از مجموع ۵۲۳ سوال دارای ضریب افتراق منفی،۲۱۷ سوال (٤۱/۵ درصد) و از ۷۱٦ سوال باقدرت افتراق ضعیف ۳۲۵ سوال (٤۵/٤ درصد) در سطح تاکسونومی یک طراحی‌شده بود. به عبارتی نزدیک به نیمی از سوالات منفی و ضعیف در سطح پایین حیطه شناختی بودند. از ٤۷۲ سوال باقدرت افتراق متوسط، ۲٤۳ سوال (۵۱/٤درصد) در سطح تاکسونومی سه و از ۳۸۹ سوال باقدرت افتراق خوب، ۱۵٦ سوال (٤۰/۱درصد) در سطح تاکسونومی سه بود. به عبارتی نزدیک به نیمی از سوالاتی که قدرت افتراق متوسط و خوب داشتند، در تاکسونومی سه‌طبقه بندی می‌شدند.تحقیق ما با مطالعه سیاح و همکاران که بر روی سوالات آزمون ارتقا دستیاری گروه اطفال در دانشگاه علوم پزشکی کاشان انجام شد، مطابقت دارد (۲۹).

در مطالعه‌ی حاضر، رابطه‌ی معناداری بین ضریب دشواری و ضریب تمایز سوالات به‌دست آمد. درجه‌ی دشواری سوال هرچه افزایش یابد سوال آسان تر بوده و لذا قدرت افتراق کمتری خواهد داشت. معمولا سوالات بسیار سخت و بسیار آسان ضریب تمایز مناسبی جهت شناسایی توانایی فراگیران نمی باشد (۳۰). Green و همکاران در تحقیق خود اعلام کردند سوالات بسیار سخت و بسیار آسان کمک کمتری به ما می‌کند تا اطلاعات مربوط به توانایی دانشجویان را به دست آوریم (۳۱) ؛ بنابراین در طراحی سوال، توجه به میزان سختی سوالات در افتراق دانشجویان موثر است.

نتیجه گیری

نتایج این پژوهش نشان داد افزایش سطح تاکسونومی سوالات می‌تواند باعث بهبود ضریب تمایز سوالات آزمون گردد و درنتیجه افتراق بهتری بین دستیاران صورت پذیرد. به‌عبارتی‌دیگر سوالاتی که تاکسونومی بالاتری دارند، قدرت افتراق بهتری نسبت به سوالات با تاکسونومی پایین تر دارند. با توجه به این‌که سوالات طراحی‌شده با تاکسونومی دو و سه، کمتر از حد استاندارد است، پیشنهاد می‌گردد اساتید در طراحی سوالات آزمون ارتقا، سوالات بیشتری با تاکسونومی بالاتر طرح نمایند.

تقدیر و تشکر

از مسوولین محترم مرکز مطالعات و توسعه‌ی دانشگاه که در اجرای این طرح پژوهشی ما را حمایت فرمودند، تقدیر می‌گردد.

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

References

1- Mc Coubrie P. Improving the fairness of multiple choice questions: a literature review. *Med teach.* 2004; 26(8): 709-12.

2- Newell FW, Leopold IH, Adler FH. The multiple choice question test of the American Academy of ophthalmology. *The American ophthal. Society Publication*. 1970; 68: 163-70.

3- Rezaii H, Shogi F, Dolati Y, Mohagheghi M, Malak zadeh R. Measurement and Evaluation in Medicine. *1 st ed Tehran: Ameh*.1998; 75. [Persian]

4- Allami A, Derakhshan F, Safdari F. Design and evaluation of multiple choice questions (MCQ).*1th ed. Qazvin.Qazvin university of Medical Science*.2011:69-70. [Persian]

5- Ware J, Vik T. Quality assurance of item writing: During the introduction of multiple choice questions in medicine for high stake examination. *Med Teach*. 2008; 29: 1-6.

6- Epstein RM. Assessment in medical education. *New England J Med*. 2007; 356(4): 387-96.

7- Tyng TY, Othman AR. The Relationship between Complexity (Taxonomy) and Difficulty. *AIP Conference Proceedings*. 2013 Apr 22, Malaysia.

8- Kaufman DM. Assessing medical students: Hit or miss, Student *BMJ*. 2001; 9: 87-8.

9- Van Der Vleuten C. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education*. 1996; 1:41- 67.

10- Mohagheghi MA, Vahidshahi K, Shakeri S. Residency test realm promote innovation and development. Ministry of Health. [citd 2013 July 20]. Available from:URL: http://cgme.behdasht.gov.ir/uploads/264_975_TarikhcheErtegha.pdf. [Persian]

11- Koshering 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): 560. [Persian]

12- Haghshenas M, Vahid Shahi K.mahmudi M,Shahbazenejod 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):120-27. [Persian]

13- Arab M. Comparative study of multiple-choice questions examinations of Medicine, Hamadan University of Medical Sciences in the second half of 1378. 4th National Congress of Medical Education. 2000: 51. [Persian]

14- Haghshenas M, Vahidshahi K, Mahmoudi M, Shahbaznejad 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): 119-26. [Persian]

15- Shakournia AH, Mozafari A, Khosravi borojeni A. Survey on Structural of MCQs of residency exam in AJUMS. *Jundishapur*

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

Scientific Medical Journal. 2009; 8(4): 491-502. [Persian]

16- Masters JC, Hulsmeyer BS, Pike ME, Leichty K, Miller MT, Verst AL. Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *J Nurs Educ*. 2001; 40(1):25-32.

17- Huxham GJ, Naeraa N. Is bloom's taxonomy reflected in the response pattern to MCQ items? *Med Educ*.1980; 14(1):6-23.

18- Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Educ Today*. 2006; 26(8): 662-71.

19- Rasoli nejad A, vakili Z,fakharian E,mosaiiebi Z, et al. Comparison tests assistants Taxonomy of kashan University of Medical Sciences 2006.Strides in Development of Medical Education. 8th National Congress of Medical Education. 2007; 217. [Persian]

20- Akhavan karbasi S. Evaluation of multiple-choice questions Pediatric Residents Yazd University of Medical Sciences. 13th National Congress of Medical Education.2012:203. [Persian]

21- Mohagheghi MA, Vahidshahi K, Shakeri S,Sabouri M,Razavi SM,Mohammadi M,et al. Compare some qualitative indicators multiple choice questions written of a specialized 86-88.Ministry of Health. [citd 2013 July 20]. Available from: URL: http://cgme.behdasht.gov.ir/uploads/264_975_Maghale_ShakhesKeifiyat.pdf . [Persian]

22- Deway RA. Writing multiple choice questions which require comprehension. Available from:.URL:http://eit.necc.mass.edu/atli/Mchoice.compreh. htm.

23- Vakili Z,sayah M, Fakharian E,Rasooli nejad SA,masibi Z,Moniri R. The psychometric properties of the multiple-choice tests residency, Kashan University of Medical. Strides in Development of Medical Education. 8th National Congress of Medical Eduction.2007; 174-75. [Persian]

24- Derakhshan F, Jahanihashemi H. Evaluation of End Term Exams of Ghazvin Medical University in 2007-2008] Special Supplement.10th National Congress of Medical Education.2009: 308-09 [Persian].

25- Basir shabestari S,sefidi F,jahani hashemi H,Taherpoor M. Compare multiple choice test, the Dental School, Qazvin University of Medical Sciences 2009-2010.12th National Congress of Medical Education.2011:335[Persian]

26- Mosavi A,sayad Z,Roozbehi A,Moosavi zade A. Analysis multiple choice question Yasooj University of Medical Sciences. 15th National Congress of Medical Education.2014:218.[Persian]

27- Meyari M, Beiglarkhani M. Quality multiple choice questions design of residency promotion Exams by feedback. *Strides in Development of Medical Education*. 2013: 10(1): 109-18.[Persian]

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

28- Vahidshahi K, Mohagheghi MA, Shakeri S,Sabouri M,Razavi SM,Mohammadi M,Ahanghari M. Comparison of some quality criteria of multiple choice questions in residency annual exam 2007 & 2008. Available From: URL: http://cgme.behdasht.gov.ir/uploads/264_781_N4_Moghayese.html. [Persian]

29- Saiah M, Momen heravi M, Vakili Z, soleimani Z, Bighdeli M, Sabori M, Motevallian SM. Prevalence of taxonomic levels, hardness test, multiple choice questions Department of Pediatrics Kashan University of Medical Sciences. 13th national conference on medical education.2012; P. 470. [Persian]

30- Hotiu A.The relationship between item difficulty and discrimination indices in multiple-choice tests in a physical science course. *Florida Atlantic University*. 2006; P; 52-4.

31- Green B, Bock R, Humphreys L, Linn R, Reckase M. Teaching guidelines for assessing computerised adaptive tests. *J Edu Meas*. 1984; 21:347–60.

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]

Quantitative and Qualitative Indicators Evaluation of Residency Exams in Qazvin University of Medical Sciences (2012-13)

Derakhshan F[1], Ahmady S[2], Allami A [3]

[1]Medical Education Faculty, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Dept. of Medical Education, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3]Dept. of Infectious Diseases, Qazvin University of Medical Sciences, Qazvin, Iran

Corresponding author: Derakhshan F, Medical Education Faculty, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Email: faribaderakhshan@gmail.com
Received: 31 Jan 2015 ***Accepted:*** 3 Aug 2015

Background and Objective: The exact design of the residency exams (standards-based) is very important. The purpose of this study was Quantitative and qualitative indicators evaluation of residency exams in Qazvin University of Medical Sciences and study of the relationships between taxonomic levels by quantitative indicators.

Materials and Methods: All questions of residency exams from 7 Medical Specialties of Qazvin University of Medical Sciences in 2012 and 2013 were studied. Taxonomy of questions in three levels: remembered, understanding and judges were determined. Discriminating power was studied in four groups as good, fair, poor, and negative. Difficulty was classified into 5 levels of easy, medium, hard, very difficult and extremely difficult. SPSS v.19 software was used for the calculation of the T-test and ANOVA.

Results: A total of 2100 questions, 38 percent of the questions reminded taxonomy, 24.1 percent of the level of understanding and 37.9 percent level were judges. 24.9 percent of questions had negative discriminating power, 34.1 percent had poor, 22.5 percent had average and 18.5 percent had good discriminating power. For the degree of difficulty, 22.1 percent simple questions, 22.6 percent average, 24.5 percent difficult, 17.5 percent very difficult and 13.2 percent was extremely difficult. A significant association was found between taxonomic questions and differentiation factor, taxonomy remembered had less differentiation index and questions, taxonomic understanding and application had differentiation index higher($p<0.0001$).

Conclusion: The results showed increased levels of taxonomic questions can enhance the discriminating power test questions. Since questions designed the taxonomy of two and three were lower than the standard (80%), teachers were recommended design more questions higher taxonomy in residency exams.

Keywords: Taxonomy, Difficulty index, Discrimination - index, Residency exam

[Downloaded from edujournal.zums.ac.ir on 2025-12-08]