

کاربرد مدل راش (روش نقشه سوال) در تعیین استاندارد و نمره برش آزمون پیش کارورزی علوم پزشکی

شیرین دخت حبیبزاده^۱، علی دلاور^{۲*}، نورعلی فرخی^۳، اصغر مینائی^۴، محمد جلیلی^۵

تاریخ پذیرش: ۹۸/۰۴/۳۱

تاریخ دریافت مقاله: ۹۸/۰۲/۰۲

چکیده

مقدمه: یکی از بحث برانگیزترین موضوعات در اندازه‌گیری‌های تربیتی تعیین نمره برش در هنگام ساخت و اجرای آزمون‌ها است. هدف این پژوهش تعیین نمره برش آزمون ملاک مرتع پیش کارورزی علوم پزشکی با استفاده از مدل راش (روش نقشه سوال) بود. روش: پژوهش حاضر از نوع پژوهش‌های توصیفی- مقطعي بود که برای این منظور، از ۱۳ عضو هیات علمی دانشگاه علوم پزشکی دانشگاه تهران، به عنوان اعضای هیئت داوران دعوت به عمل آمد. ابزار این پژوهش، آزمون چندگرینه‌ای ۲۰۰ سوالی پیش کارورزی (پره انترنی) قطب ده علوم پزشکی کشور در شهریور ماه ۱۳۹۵ بود که تعداد ۲۷۴ داوطلب از دانشجویان دانشگاه‌های علوم پزشکی به آن پاسخ داده بودند. در فرآیند تعیین استاندارد و نمره برش؛ یک نمره برش برای تفکیک داوطلبان در سطوح عملکردی قبول/ رد تعیین شد. داده‌ها با استفاده از نرم‌افزار آماری Winstep تحلیل گردید.

یافته‌ها: نتایج نشان داد که این آزمون از اعتبار سوال $IR=0.93$ و اعتبار فرد $PR=0.98$ برخوردار است که نشان دهنده مناسب بودن واریانس نمونه، طول تست و همچنین سوالات از حوزه‌های شانزده‌گانه به تناسب انتخاب گردیده است. اجرای پنل داوران و بررسی سوالات با روش نقشه سوال نشان داد که توافق داوران پس از سه مرحله داوری شامل پارامتر دشواری برابر (0.48) و نمره خام برابر 123 و معادل پارامتر توانایی (0.58) بوده است.

نتیجه‌گیری: بررسی روایی نمره برش نشان داد که استفاده از روش نقشه سوال توانسته است توافق بالایی بین داوران نسبت به تعیین نمره برش ایجاد نماید. نتیجه نشان داد که هیات داوران رضایت بالایی از این فرآیند از کاربردی بودن و اجرایی بودن مراحل اجرا دارد.

کلیدواژه‌ها: نمره برش، مدل راش، نقشه سوال، پیش کارورزی، آزمون

مقدمه

توانایی‌های معین و مورد انتظار دست یافته‌اند. به طور کلی برای دستیابی به این گواهی‌نامه‌ها، آمیزه‌های از دانش‌ها و مهارت‌های لازم و کسب نمره قبولی در آزمون ضروریست. نمره قبولی در آزمون یکی از الزامات مهم و بحث برانگیز جهت حفظ ارزش و اعتبار آزمون‌های ملاک مرتع

(Criterion_Reference) محسوب می‌شود. فرایند تعیین حداقل توانمندی و شایستگی‌های لازم داوطلبان براساس نوع دانش و میزان آن تعیین استاندارد (Standard Setting) و تبدیل آن به نمره‌ای واقعی و قابل اندازه‌گیری شامل حداقل نمره قبولی، را نمره برش (Cut Score) می‌نامند (۳). سیزک (Cizek) فرآیند تعیین استاندارد را «پیگیری درست سیستم تجویزی و منطقی قوانین یا رویه‌هایی که به تعیین عددی منتهی می‌شوند که دو یا چند سطح یا درجات عملکردی را از هم جدا می‌کند» تعریف می‌نماید (۴). در

هدف آموزش مبتنی بر کسب شایستگی حرفه‌ای، افزایش دانش‌ها و مهارت‌ها و نگرش‌های دانشجویان در رشته پزشکی می‌باشد. آموزش مبتنی بر کسب شایستگی حرفه‌ای به دانشجویان کمک می‌نماید تا بتوانند خدمات مناسب حرفه‌ای به جامعه ارائه نمایند (۱). اعطای گواهی‌نامه‌ی حرفه‌ای و تخصصی به داوطلبان، نشان می‌دهد که داوطلبان پذیرفته شده در حوزه تخصصی خود به سطح دانش، مهارت یا

نویسنده مسئول: علی دلاور، گروه سنجش اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران
delavarali@yahoo.com
شیرین دخت حبیبزاده (دانشجوی دکتری سنجش و اندازه‌گیری)، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران
نورعلی فرخی، گروه سنجش اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران
اصغر مینائی، گروه سنجش اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران
محمد جلیلی، گروه طب اورژانس و گروه آموزش پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران

گردد هم می‌آیند تا داوطلب با حداقل شایستگی را براساس دانش یا توانایی مورد قبول مشخص نمایند. در این میان یک تسهیل کننده و راهنمای نیز روند مباحثه میان اعضای هیات مشاوره (این افراد فهرستی از علوم، مهارت و توانایی‌هایی را ارایه می‌کنند که برای دستیابی به نمره قبولی باید وجود داشته باشند) را میسر می‌سازد (۴، ۳).

برای تبدیل استاندارد قبولی به حداقل نمره، رویکردهای متعددی وجود دارد. در روش‌های آنگوف (Angoff) و ابیل (Ebel) (کار برآورده تعداد آزمودنی‌های با حداقل توانایی (Minimally Competent Candidate) که می‌توانند یک سوال را درست پاسخ دهند بر عهده اعضای هیات داوران است. روش نیدلسکی (Nedelsky) اعضاً هیات داوران را ملزم می‌سازد تا عوامل خطا را شناسایی کنند که ممکن است سبب شوند (MCC) سوال مربوطه را درست پاسخ ندهند (۱۵، ۱۴). زمانی که داده‌های اعضاً هیات داوران با هم مطابقت داده می‌شوند، پس از آرایه برخی از اطلاعات دور دوم قضاوت درمورد سوال‌ها آغاز می‌شود. این اطلاعات اغلب عبارتند از داده اثر (Impact Data) حاصل از توافق هیات داوران (۱۶، ۱۷). گاهی از روش تعیین استانداردی استفاده می‌شود که در آن برای تعیین نمره برش از اطلاعات هنجاری استفاده می‌گردد (۱۸، ۱۹).

یکی از رایج‌ترین روش‌های تعیین نمره برش در آزمون‌های ملاک مرجع روش بوک مارک (Bookmark) می‌باشد (۲۰-۲۲). روش بوک مارک یک رویکرد نسبتاً جدید است که در سال ۱۹۹۶ توسط گروه تحقیقاتی CTB/ McGraw-Hill برای استفاده در آموزش K-12 ساخته شده است (۱۸، ۲۳). روش نقشه سوال (Item- Mappig) برگرفته از روش بوک مارک می‌باشد، این روش از نظریه سوال- پاسخ براساس مدل راش (Item- Response Theory: Rasch Model) فرآیند تعیین استاندارد و نمره برش استفاده می‌نماید. این روش جایگزین مناسبی برای روش آنگوف است که تاکنون در آزمون‌های چندگزینه‌ای و مخصوصاً آزمون‌هایی که با هدف اعطاء گواهی‌نامه اجرا می‌شوند، مورد استفاده قرار می‌گرفت (۸).

پرکاربردترین روش تعیین نمره برش برای آزمون‌های چند گزینه‌ای و آزمون‌های اعطای گواهی‌نامه، روش آنگوف و آنگوف اصلاح شده می‌باشد (۲۴). در این روش از داوران خواسته می‌شود تا نسبت داوطلبانی با حداقل صلاحیت که به سوال پاسخ درست می‌دهند را برآورد نمایند (۲۵). انتقاد عمده به

آزمون‌های ارایه گواهی، تعیین منطقی استاندارد امری ضروریست به عبارت دیگر، استاندارد قبولی باید حداقل میزان شایستگی لازم جهت دریافت گواهی‌نامه را مشخص نماید (۵، ۶). برنامه دوره دکتری عمومی شامل دروس عمومی و علوم پایه، نشانه شناسی و فیزیوپاتولوژی، کارآموزی و کارورزی می‌باشد. به دلیل حساسیت بالای این رشته و تاثیر مستقیم توانمندی‌های حرفه‌ای دانشجویان این رشته بر سلامت جسمانی و روانی فرد فرد افراد جامعه در طول دوران تحصیل، دانشجویان رشته‌های پزشکی باید بتوانند از سه مقطع و سه آزمون عبور نمایند. پس از سپری شدن دوره عمومی و علوم پایه، دوره نشانه شناسی و فیزیوپاتولوژی به عنوان اولین مرحله آموزش بالینی در دوره دکتری عمومی آغاز و آزمون جامع پیش کارورزی در پایان دوره کارآموزی برگزار می‌گردد. در صورتی که دانشجو بتواند آزمون پیش کارورزی را با موفقیت طی نماید می‌تواند وارد دوره کارورزی گردد (۷). نکته بسیار مهم در تعیین استاندارد این آزمون تعیین نمره برش به گونه‌ای است که بتواند استاندارد عملکرد مناسبی را نشان دهد. این نمره باید بتواند بین داوطلبانی که به سطح شایستگی دست یافته‌اند و داوطلبانی که نتوانسته‌اند شایستگی حرفه‌ای لازم را کسب نمایند تمایز قائل شوند (۸). فرآیند تعیین استاندارد در دو مرحله رخ می‌دهد. در مرحله نخست مفهوم استاندارد عملکرد تعریف می‌شود. سپس این استاندارد به حداقل نمره یا نمره قبولی تبدیل می‌گردد (۳، ۸). روش‌های تعیین استاندارد و نمره برش مناسب با اهداف آزمون‌ها و نوع آزمون‌ها متنوع هستند. در روش‌های آزمودنی محور (Candidate- Centered) فرایند داوری براساس بحث و فضای عملکرد فرد در آزمون صورت می‌گیرد و در روش‌های آزمون محور (Test- Centered) تمرکز بر سوال‌های آزمون می‌باشد (۹-۱۱). همچنین در روش‌های هنجار- مرجع (Norm_Refrenced) مبنای تعیین نمره برش، تعداد یا درصد افراد از پیش تعیین شده برای قبولی بوده ولی در روش‌های ملاک مرجع مبنای تعیین نمره برش، رعایت معیارها و ملاک‌هایی است که از نوع عملکرد سایر آزمودنی‌ها تاثیر نمی‌گیرد (به عبارت دیگر هر فردی که داشت لازم را داشته باشد می‌تواند در این آزمون قبول شود). از روش‌های ملاک مرجع و آزمون محور غالباً در آزمون سرنوشت ساز استفاده می‌شود (۱۱-۱۳).

در روش‌های آزمون محور و ملاک مرجع ابتدا گروهی از کارشناسان موضوعی (داوران، ارزیابان یا اعضای هیات مشاوره)

درستی بتواند آزمودنی‌هایی که شایستگی‌های لازم را کسب نموده‌اند را از سایر آزمودنی‌ها، تفکیک نماید از اهمیت بالایی برخوردار است. تعیین حد نصاب قبولی یا نمره برش در آزمون پیش‌کارورزی به شیوه‌ای کنونی که داوطلبانی در آزمون قبول اعلام می‌گردند که بتوانند ۷۰ درصد میانگین نمرات ۵ درصد افرادی که بالاترین نمره را در سطح کشور کسب نموده‌اند؛ را کسب نمایند^(۷)، برای آزمون‌های ملاک مرجع مناسب نبوده است و بیشتر با هدف مقایسه افراد با یکدیگر کاربرد دارد. لذا با توجه به مزیت‌هایی که روش نقشه سوال دارد، در این پژوهش به تعیین استاندارد و نمره برش آزمون براساس روش نقشه سوال پرداخته شد. تحقیق حاضر بر روش تعیین استاندارد و نمره برشی تمرکز دارد که برای آزمون‌های چهارگزینه‌ای ملاک مرجع و سرنوشت‌ساز همانند آزمون‌های جامع علوم پزشکی مناسب است. این پژوهش به بررسی این سوال می‌پردازد که نمره برش مناسب با استفاده از روش نقشه سوال در آزمون جامع پیش‌کارورزی چقدر است؟ و آیا این روش از روایی رویه‌ای مناسبی برخوردار است؟

روش‌ها

پژوهش حاضر در زمرة مطالعات توصیفی- مقطوعی است. جامعه آماری پژوهش شامل کلیه داوطلبان آزمون پیش‌کارورزی (پره انترنی) شرکت‌کننده در آزمون‌های جامع پزشکی شهریور ماه ۱۳۹۵ به تعداد ۲۷۴ نفر^(۱۵) زن، ۱۲۳ نفر مرد بود. پژوهش حاضر به بررسی نمره برش آزمون جامع پیش‌کارورزی (پره انترنی) قطب دو کشوری می‌پردازد. این آزمون هر سال در دو نوبت شهریور و اسفند ماه در رشته‌های پزشکی دانشگاه‌های علوم پزشکی تهران، شهید بهشتی، ایران، بقیه‌الله، ارتش، بهزیستی و توانبخشی اجرا می‌گردد. آزمون شامل ۲۰۰ سوال چهار گزینه‌ای با نمره‌گذاری صحیح ۱ و غلط ۰ در موضوعات داخلی، جراحی، کودکان، زنان، مغز و اعصاب، عفونی، رادیولوژی، آسیب‌شناسی، روان‌پزشکی، پوست، جراحی استخوان و مفاصل، جراحی کلیه و مجاری ادراری تناسلی، چشم پزشکی، گوش و حلق و بینی، آمار و اپیدمیولوژی، فارماکولوژی و اخلاق پزشکی است. در پژوهش حاضر با توجه به بکارگیری نظریه سوال-پاسخ (نقشه سوال) در برآورد نمره برش آزمون جامع پیش‌کارورزی پزشکی، دستیابی به برآورد پارامترهای سوال و توانایی تنها با استفاده از روش‌های کمی امکان‌پذیر است. لذا در این پژوهش، ابتدا با استفاده از نرم‌افزار Winstep 3.92.1

روش آنگوف این است که داوران برآورد صحیحی از دشواری سوالات و نحوه ارتباط آن با داوطلب با حداقل صلاحیت ندارند و برآوردها در بسیاری مواقع ناهمسو و متناقض می‌باشد (Cizek, ۲۶). سیزک (Wang) به نقل از وانگ^(۲۷) بیان نمود که داوران در تعیین نمره برش به روش آنگوف، قادر به پیش‌بینی دقیق عملکرد آزمودنی‌های مرزی در سوال‌های آزمون نیستند. روش نقشه سوال با استفاده از مدل راش روند پیش‌بینی را ساده می‌کند و کمبود شناختی روش‌های دیگر را برطرف می‌نماید از سویی دیگر در روش آنگوف داوران به بررسی عملکرد داوطلبان در تک تک سوالات می‌پردازند در حالی که روش نقشه سوال یک تصویر کلی از همه سوالات براساس پارامتر دشواری برآورده شده را در هیستوگرام ارائه می‌نماید. به این ترتیب فرایند تصمیم‌گیری برای داوران تسهیل می‌گردد. داوران از طریق روش نقشه سوال می‌توانند سوال‌ها را در یک هیستوگرام مشاهده کنند. این سوال‌ها به ترتیب پارامتر دشواری سوال (Rasch Item Difficulty Parameter) دشواری کم شروع و به سمت سوالات با پارامتر دشواری بالا در روی یک هیستوگرام قرار گرفته‌اند. در این روش هنگامی که توانایی داوطلب با دشواری سوال برابر باشد احتمال پاسخ صحیح به سوال ۵۰ درصد است و این مزیت عدمه روش نقشه سوال می‌باشد که کار برآورده را برای داوران آسان می‌نماید. داوران سوال‌ها را هم زمان مرور می‌کند، اعضاء پس از بررسی هر سوال، مشخص می‌کنند که آیا به احتمال ۵۰ درصد (یا بیشتر) داوطلبان مرزی می‌توانند به سوال مربوطه پاسخ درست بدeneند یا خیر؟ این فرآیند آنقدر تکرار می‌شود تا اعضاء به این اتفاق نظر برسند که در چه سوال یا سوال‌هایی داوطلب با حداقل توانایی با احتمال ۵۰ درصد نمی‌توانند به سوال یا سوال‌هایی بعدی پاسخ درست بدeneند. موقعیت سوال یا سوال‌ها در هیستوگرام با سطح توانایی معینی در انتبطاق است سپس این سطح توانایی به حداقل نمره در آزمون تبدیل می‌گردد که نمره برش آزمون نامیده می‌شود^(۱۸). آزمون‌های پیش‌کارورزی علوم پزشکی با هدف تصمیم‌گیری درمورد ارتقاء دانشجویان به مقطع بالاتر تحصیلی طراحی و اجرا می‌گردد، دوره‌ی انترنی آخرین دوره‌ی آموزش پزشکی عمومی می‌باشد و دانشجویان باید بتوانند با کسب شایستگی‌های لازم در زمینه دانش و مهارت‌های پزشکی به این دوره وارد شوند (۲۸). به همین دلیل سنجش توانایی‌های مورد نیاز جهت ورود به دوره‌ی انترنی و به دنبال آن، تعیین نمره‌ای در آزمون که به

این آماره به حجم نمونه بسیار حساس است. برای نمونه‌های با حجم کوچک‌تر از ۹۰ نفر هر نوع داده‌هایی با مدل برازش می‌باید، در حالی که برای نمونه‌های بزرگ‌تر از ۹۰۰ نفر هیچ نوع داده‌ای با مدل برازش ندارد (۳۰). که با توجه به جدول شماره یک مقدار مشاهده شده این آماره در هر دو حالت Outfit و Infit) برابر صفر است. لذا با توجه به مقادیر مشاهده شده هر یک از این آماره‌ها می‌توان گفت که سوالات آزمون پیش‌کارورزی با مدل تکپارامتری راش (نظریه سوال-پاسخ) برازش دارند. پس از بررسی شاخص‌های برازش آزمون و محاسبه پارامترهای دشواری و توانایی سوالات براساس مدل راش، با استفاده از روش‌های کیفی و بهره‌گیری از نظر متخصصان، نمره برش سوالات مشخص شد. بدین منظور با توجه به این که اجرای آزمون جامع پیش‌کارورزی در شهریور ماه ۱۳۹۵ بر عهده دانشگاه علوم پزشکی تهران بوده است، لذا برای بررسی نمره برش، لیستی شامل ۱۳ نفر از اعضاء هیات علمی دانشگاه علوم پزشکی تهران شناسایی، انتخاب و دعوت به همکاری شدند.

در انتخاب آن‌ها تلاش شد تا بیشترین تنوع از نظر تخصص، سابقه آموزشی، سن در نظر گرفته شود. جلسه تعیین استاندارد به صورت کارگاهی نیم‌روزه (۶ ساعت) برگزار شد. با توجه به ملاحظات اخلاقی درمورد شرکت کنندگان در پژوهش، آن‌ها از هدف تحقیق مطلع شده و به طور داوطلبانه و اختیاری در پژوهش شرکت داشتند و تمام اطلاعات محترمانه می‌باشد. در ابتدای جلسه داوران با مفاهیم تعیین استاندارد و نمره برش و اهمیت آن در آزمون‌های پزشکی آشنا شدند، سپس به بررسی توانایی‌های داوطلب با حداقل شایستگی پرداختند، پس از آن سوالات آزمون جامع پره انتزنا که با استفاده از نرم‌افزار تحلیل شده بودند، برحسب پارامتر دشواری از آسان‌ترین سوال تا دشوارترین سوال در قالب دفترچه جدیدی از سوالات آزمون تنظیم شد و در اختیار داوران قرار گرفت. در گام بعدی داوران به این سوال پاسخ دادند که «کدام صفحه و کدام سوال است که داوطلب با حداقل صلاحیت می‌تواند به آن پاسخ داده و به سوالات بعد از آن نمی‌تواند پاسخ درست دهد؟» داوران به بررسی سوالات پرداخته و شماره سوال و صفحه سوال را در کاربرگی که در اختیارشان قرار گرفت وارد نمودند. در شکل شماره یک نمونه‌ای از سوالات آزمون دفترچه جدید ارائه شده است.

سوالات آزمون به شیوه کمی و با استفاده از مدل راش مورد تجزیه و تحلیل قرار گرفت.

ضریب اعتبار سوالات آزمون (Item ReliabilityCofficience) جامع پیش‌کارورزی پزشکی برابر ۰/۹۸ براورد شد که نشان دهنده این مطلب است که میزان خطأ در پاسخ به سوالات بسیار کم بوده است. همچنین ضریب اعتبار شخصی (Person Reliability Cofficience) برابر ۰/۹۳ بود که نشان می‌دهد که واریانس نمونه و طول آزمون مناسب است، از انواع سوالات حوزه‌های شانزده‌گانه به اندازه‌ی کافی نمونه سوال در آزمون استفاده و اینکه سوالات متناسب با سطح کاندیداها انتخاب شده، توانسته بین افراد تمایز ایجاد کند (۲۹). شاخص برازش شخصی (Person Sepration Index) نیز برابر ۳/۵۲ براورد شد که نشان دهنده توانایی آزمون در برش و تمیز آزمودنی‌ها است. همچنین ضریب برازش سوالات (Item sepration Index) ۷/۷۲ براورد گردیده که نشان می‌دهد این آزمون در تمیز سوالات خردۀ آزمون‌های شانزده‌گانه توانمند عمل کرده است. همچنین ضریب آلفای آزمون برابر ۰/۹۳ بدست آمد که نشان دهنده پایایی مطلوب آزمون است.

جدول ۱: آماره‌های zstd و mnsq جامع پیش‌کارورزی پزشکی

آماره	Outfit		Infit	
	Mnsq	Zstd	Mnsq	Zstd
میانگین	۱	۰	۱	۰
انحراف استاندارد	۰/۲۲	۱/۶	۰/۱۱	۱/۵
ماکزیمم	۱/۳۹	۵/۰	۲/۱۵	۶/۳
مینیمم	۷۷/۰	-۳/۴	۰/۶۳	-۳/۱

در بررسی برازش سوالات آزمون با استفاده از نرم‌افزار winsteps باید به دو آماره Mnsq و Zstd توجه داشت. آماره Mnsq در دامنه‌ای از صفر تا ۱ نهایت است که جهت برازش سوالات آزمون این آماره باید یک باشد. مقادیر بالاتر از یک بیانگر انحراف از تک بعدی بودن و مقادیر پایین‌تر از نیز بیانگر بیش برازش الگوهای پاسخ با مدل راش است و این به معنای وجود وابستگی در میان پاسخ‌ها یا سوالات می‌باشد. همان‌طور که در جدول شماره یک مشاهده می‌شود مقدار این آماره در Outfit و Infit برابر یک است. همچنین آماره Zstd نیز آماره معنی‌داری آماری انحراف داده‌ها از مدل راش را آزمون می‌کند. این آماره از تبدیل آماره Mnsq به آماره t به دست می‌آید در نمونه‌های بزرگ دارای توزیع Z می‌باشد، این آماره معنی‌داری انحراف داده‌ها از مدل راش را آزمون می‌کند.

بحث گذاشته شد.

در مرحله سوم (راند سوم داوری)، علاوه بر دفترچه سوالات و جدول جایگاه سوالات برحسب پارامتر دشواری آن‌ها، نمودار هیستوگرام سوالات، در اختیار داوران قرار گرفت. در این نمودار، سوالاتی که دارای ضربه دشواری تقریباً یکسانی هستند، بروی یک محور قر داشت و سپس مجدداً از داوران درخواست گردید که ستوانی از سوالات را بروی هیستوگرام انتخاب نمایند که از نظر آنان داوطلب با حداقل صلاحیت به سوالات قبل آن به درستی پاسخ می‌دهد و به سوالات بعد از آن نمی‌تواند به درستی پاسخ دهد. در این مرحله داوران با هم فکری با یکدیگر به بررسی ستون‌های هیستوگرام پرداخته و سپس یک ستون را انتخاب نمودند. در مرحله سوم نیز راهنمای جلسه پاسخ داوران را ثبت نمود. جهت بررسی روایی تعیین نمره برش از روایی رویه‌ای استفاده گردید. روایی رویه‌ای از طریق شواهدی تعیین گردید که مراحل و فرآیند تعیین استاندارد و نمره برش را در بر می‌گیرد. به این منظور پس از جلسه‌ی تعیین نمره برش و به جهت ارزیابی فرایند اجرا شده، داوران آزمون پیش‌کارورزی در قالب یک فرم ارزیابی به سوالات در دامنه‌ای از کاملاً موافق تا کاملاً مخالف همانند جدول پاسخ دادند.

یافته‌ها

پس از دریافت پاسخ داوطلبان آزمون پیش‌کارورزی شهریور ۱۳۹۵ از دانشگاه علوم پزشکی تهران، داده‌ها توسط نرم‌افزار winstep تحلیل و پارامتر دشواری و پارامتر توانایی سوالات محاسبه گردید. آسان‌ترین سوال، سوال شماره ۱۰۲ با پارامتر دشواری ۳/۵۲ و دشوارترین سوال، شماره ۵۳ با پارامتر دشواری ۳/۶۳ بود. در مرحله اول داوری، داوران دفترچه سوالات را که به ترتیب پارامتر دشواری سوالات از آسان به دشوار مرتب شده بود، دریافت نمودند و نمره برش مورد نظر خود را اعلام کردند. نتایج مرحله اول داوری در جدول شماره دو آمده است. در راند دوم داوری از اعضای هیئت علمی خواسته شد تا با در اختیار داشتن پارامتر دشواری محاسبه شده برای هر سوال که در جدول نشان داده شده است، سوالی را که از نظر آنان داوطلب با حداقل صلاحیت، قادر به پاسخ‌گویی به سوالات بعد از آن نیست، مشخص کنند.

شماره صفحه:	سوال شماره:
۲۸	۶۶
	- مرد ۳۵ ساله به دنبال تصادف با توبیل، به اورژانس ارجاع شده است. در معاینه، کاهش سطح هوشیاری (GCS:9/15)، اکسیوز امراض مرد چشم، ورید روزگر لارچیه داشته و هیچین تندرنس قفسه سینه و شکم و علامت کبودی، خراشیدگی و آشفتگی سرتاسری نداشت. فشار خون وی ۷۵/۵۵ mmHg و تعداد ضربان قلب ۱۵۰ و تعداد تنفس ۳۲/min است. کدام اقام اولویت پیشتری دارد؟
	(الف) پرکاربوکسٹر
	(ب) گلداشتن لوله قفسه صدری
	(ج) سستی اورژانسی مفز
	(د) مشاوره فوری جراحی برای انجام لایارانوئی اورژانس
کلید: ب	

شکل ۱: سوالات دفترچه آزمون جامع پیش‌کارورزی برحسب پارامتر دشواری

همانطور که در شکل شماره یک مشاهده می‌شود، دفترچه سوالات آزمون شامل: شماره سوال در دفترچه آزمون جامع پیش‌کارورزی، شماره صفحه سوال در دفترچه جدید آزمون جامع پیش‌کارورزی (قطب ۱۰ کشوری) که برحسب دشواری سوالات مرتب شده‌اند، متن سوال، گزینه‌ها و کلید آن ارائه شده است. داوران در طی سه مرحله به بررسی و انتخاب نمره برش آزمون با روش نقشه سوال پرداختند. در مرحله اول مدل راش مرتب شده بود در اختیار داوران قرار گرفت و داوران هر یک از سوالات را از ابتدای دفترچه مطالعه نموده و سپس نمره برش آزمون را در صفحه سوالی قرار دادند که از نظر آن‌ها داوطلب با حداقل صلاحیت قادر به پاسخ‌گویی به سوالات بعد از آن نیست. شماره سوال و شماره صفحه دفترچه سوال انتخابی توسط راهنمای جلسه ثبت گردید و سپس انتخاب داوران در جلسه مورد بحث قرار گرفت.

در مرحله بعدی (راند دوم داوری)، جدول سوالات آزمون جامع پیش‌کارورزی پزشکی که برحسب پارامتر دشواری مدل راش از ساده به دشوار مرتب شده و در اختیار داوران قرار گرفت و داوران علاوه بر دفترچه سوالات، جدول مرتب شده سوالات همراه با ویژگی پارامتر دشواری سوالات را دریافت نمودند. داوران مجدداً سوالی را انتخاب نمودند که از نظر آنان داوطلب با حداقل صلاحیت به سوالات تا قبل از آن به درستی پاسخ داده و به سوالات بعد از آن نمی‌تواند پاسخ درست دهد. نتایج دور دوم داوری نیز برحسب شماره سوال و شماره صفحه انتخابی توسط هر کدام از داوران، توسط راهنمای جلسه ثبت گردید و سپس دلایل انتخاب سوال مربوطه در گروه داوران به

جدول ۲: نقطه برش از منظر داوران در مرحله اول و دوم داوری

داور شماره	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳
مرحله اول	۱۲۷	۱۲۷	۶۳	۴۶	۸۳	۱۸۶	۹۴	۱۳۷	۶	۴۵	۹۲	۹۲	۸
مرحله دو	۱۲۷	۱۲۷	۷۱	۳۶	۳۸	۱۸۶	۹۳	۱۳۷	۶	۵۲	۹۲	۴۰	۸

(استاندارد) مشخص کردند. در مرحله دوم، در کنار دفترچه سوالات جدولی حاوی پارامتر دشواری محاسبه شده براساس مدل راش در اختیار داوران قرار گرفت. در این جدول سوالات براساس پارامتر دشواری مدل تک پارامتری راش از ساده به دشوار مرتب شده بودند و مجدداً سوال اولیه از داوران پرسیده و داوران سوال مورد نظر را انتخاب نمودند.

همانطور که در جدول شماره دو مشاهده می‌شود، هر یک از ۱۳ داور، در مرحله اول داوری در پاسخ به سوال «کدام صفحه و کدام سوال است که داوطلب با حداقل صلاحیت می‌تواند به آن پاسخ داده و به سوالات بعد از آن نمی‌تواند به درستی پاسخ دهد؟ به گونه‌ای متفاوت عمل نموده‌اند و هر کدام از داوران سوالی را متفاوت از دیگری، به عنوان نقطه برش

جدول ۳: مشخصات سوالات آزمون پیش کارورزی براساس مدل کلاسیک و مدل راش (IRT)

درجه دشواری (IRT)	درجه دشواری (IRT) بر حسب درصد)	کلاسیک برش سوال	شماره سوال	شماره صفحه	درجه دشواری (IRT)	درجه دشواری (IRT) بر حسب درصد)	کلاسیک برش سوال	شماره سوال	شماره صفحه	درجه دشواری (IRT)	درجه دشواری (IRT) بر حسب درصد)	کلاسیک برش سوال	شماره سوال	شماره صفحه	
-0/64	72/35	64	126	-1/29	83/57	147	26	-1/24	82/84	41	101	-3/52	97/81	102	1
-0/64	72/35	16	127	-1/29	83/57	59	27	-1/21	82/48	111	102	-2/5	96/62	72	2
-0/64	72/35	15	128	-1/29	83/57	54	28	-1/16	51/75	172	103	-2/71	95/25	107	2
-0/64	72/35	13	129	-1/26	83/21	144	29	-1/16	81/75	157	104	-2/49	94/16	5	4
-0/6	72/62	158	130	-1/24	82/84	149	30	-1/16	87/51	104	105	-2/42	93/79	36	5
-0/25	65/69	43	131	-0/58	72/26	74	31	-1/16	81/75	99	106	-2/24	92/7	197	6
-0/23	65/32	169	132	-0/58	72/26	71	32	-1/16	81/75	48	107	-2/39	91/6	148	7
-0/21	64/96	136	133	-0/58	72/26	49	33	-1/06	80/29	66	108	-2/09	91/6	88	8
-0/21	64/96	98	134	-0/56	71/89	193	34	-1/06	80/29	55	109	-1/9	90/14	106	9
-0/19	64/59	22	135	-0/56	71/89	146	35	-0/98	79/19	62	110	-1/86	89/78	163	10
-0/14	63/5	125	136	-0/56	71/89	65	36	-0/98	79/19	47	111	-1/86	89/78	154	11
-0/14	63/5	91	137	-0/54	71/53	81	37	-0/91	78/1	118	112	-1/82	89/41	171	12
-0/14	63/5	8	138	-0/54	75/31	4	38	-0/91	78/1	9	113	-1/78	89/8	122	13
-0/06	61/67	188	139	-0/52	71/16	116	39	-0/89	77/73	61	114	-1/71	88/32	133	14
-0/04	61/31	115	140	-0/52	71/16	86	40	-0/87	77/73	83	115	-1/71	88/32	103	15
-0/04	61/31	89	141	-0/52	71/16	57	41	-0/87	77/73	6	116	-1/71	88/32	46	16
-0/02	60/94	137	142	-0/52	71/16	44	42	-0/76	75/54	50	117	-1/63	87/59	113	17
0/01	60/21	141	143	-0/52	71/16	7	43	-0/72	74/81	153	118	-1/67	86/86	120	18
0/03	59/85	159	144	-0/5	70/80	100	44	-0/72	74/81	124	119	-1/57	86/86	105	19
0/03	59/85	114	145	-0/5	70/8	38	45	-0/72	74/81	92	120	-1/53	86/49	196	20
0/03	59/85	30	146	-0/5	70/8	12	46	-0/72	74/81	69	121	-1/53	86/49	95	21
0/04	59/48	155	147	-0/46	70/8	110	47	-0/68	74/08	167	122	-1/53	86/49	93	22
0/08	58/75	143	148	-0/46	70/8	40	48	-0/68	74/08	127	123	-1/5	86/13	51	23
0/08	58/75	84	149	-0/44	69/7	73	49	-0/66	73/72	178	124	-1/47	85/76	22	24
0/08	58/75	25	150	-0/44	69/7	42	50	-0/66	73/72	134	125	-1/38	84/87	35	25

نرجه دشواری (IRT)	نرجه دشواری حسب درصد (کلاسیک)	شماره سوال	شماره صفحه	نرجه دشواری (IRT)	نرجه دشواری حسب درصد (کلاسیک)	شماره سوال	شماره صفحه	نرجه دشواری (IRT)	نرجه دشواری حسب درصد (کلاسیک)	شماره سوال	شماره صفحه	نرجه دشواری (IRT)	نرجه دشواری حسب درصد (کلاسیک)	شماره سوال	شماره صفحه
۱/۲	۳۳/۵۷	۱۶۱	۱۷۶	۰/۴	۴۵/۵۱	۱۱۲	۷۶	۰/۰۹	۵۸/۳۹	۱۲۸	۱۵۱	-۰/۴۳	۶۹/۳۴	۱۶۸	۵۱
۱/۲۲	۳۳/۲۱	۱۳۱	۱۷۷	-۰/۴۶	۰/۵	۱۹۸	۷۷	-۰/۱۲	۵۷/۶۶	۱۸۶	۱۵۲	-۰/۴۳	۶۹/۳۴	۷۶	۵۲
۱/۲۴	۳۲/۸۴	۲۱	۱۷۸	-۰/۴۶	۰/۵	۱۸۵	۷۸	-۰/۱۲	۵۷/۶۶	۹۰	۱۵۳	-۰/۴۱	۶۸/۹۷	۱۴۲	۵۳
۱/۲۶	۳۲/۴۸	۱۳۲	۱۷۹	-۰/۴۶	۰/۵	۳۹	۷۹	-۰/۱۴	۵۷/۲۹	۶۷	۱۵۴	-۰/۴۱	۶۸/۹۷	۹۴	۵۴
۱/۲۹	۳۱/۷۵	۱۸۹	۱۸۰	-۰/۴۹	۴۹/۲۷	۱۸	۸۰	-۰/۱۷	۵۶/۵۶	۱۹۱	۱۵۵	-۰/۳۷	۶۸/۳۴	۷۰	۵۵
۱/۳۱	۳۱/۳۸	۲۷	۱۸۱	-۰/۵۷	۴۷/۴۴	۱۷۰	۸۱	-۰/۱۷	۵۶/۵۶	۱۵۰	۱۵۶	-۰/۳۵	۶۷/۸۸	۱۷	۵۶
۱/۳۵	۳۰/۶۵	۳۱	۱۸۲	-۰/۵۷	۴۷/۴۴	۹۶	۸۲	-۰/۱۹	۵۶/۲	۱۳۰	۱۵۷	-۰/۲۸	۶۶/۴۲	۱۰	۵۷
۱/۴	۲۹/۶۵	۱۰۱	۱۸۳	-۰/۶۲	۴۶/۳۵	۳۳	۸۳	-۰/۱۹	۵۶/۲	۱۱۹	۱۵۸	-۰/۲۶	۶۰/۰۵	۱۶۶	۵۸
۱/۴۲	۲۹/۱۹	۷۵	۱۸۴	-۰/۶۵	۴۵/۶۲	۱۵۱	۸۴	-۰/۲۱	۵۵/۸۳	۱۳۵	۱۵۹	-۰/۲۶	۶۰/۰۵	۱۲۶	۵۹
۱/۴۴	۲۸/۸۳	۱۹	۱۸۵	-۰/۶۷	۴۵/۲۵	۱۰۸	۸۵	-۰/۲۲	۵۵/۴۷	۱۵۲	۱۶۰	-۰/۲۶	۶۰/۰۵	۲۹	۶۰
۱/۴۶	۲۸/۴۶	۸۷	۱۸۶	-۰/۷	۴۵/۵۲	۱۷۶	۸۶	-۰/۱۸	۴۰/۵۱	۵۶	۱۶۱	-۰/۲۴	۵۵/۱	۵۲	۶۱
۱/۴۸	۲۸/۱	۱۸۲	۱۸۷	-۰/۷۲	۴۴/۱۶	۱۴	۸۷	-۰/۹۱	۳۹/۷۸	۱۹۲	۱۶۲	-۰/۲۴	۵۵/۱	۴۵	۶۲
۱/۵	۲۷/۷۳	۱۲۱	۱۸۸	-۰/۷۷	۴۳/۰۶	۶۰	۸۸	-۰/۹۱	۳۹/۷۸	۵۸	۱۶۳	-۰/۲۴	۵۵/۱	۳۴	۶۳
۱/۵۶	۲۶/۶۴	۳۷	۱۸۹	-۰/۷۸	۴۲/۷	۷۹	۸۹	-۰/۹۵	۳۹/۰۵	۱۶۵	۱۶۴	-۰/۲۵	۵۴/۷۴	۱۳۹	۶۴
۱/۶	۲۵/۹۱	۲۳	۱۹۰	-۰/۸۵	۴۱/۲۴	۱	۹۰	-۰/۹۵	۳۹/۰۵	۹۷	۱۶۵	-۰/۲۵	۵۴/۷۴	۸۲	۶۵
۲/۱۳	۱۷/۵۱	۲	۱۹۱	۱/۶۴	۲۵/۱۸	۱۶۴	۹۱	۱/۱۸	۳۶/۱۳	۱۱۷	۱۶۶	-۰/۳۲	۵۴/۷۴	۱۸۴	۶۶
۲/۱۸	۱۶/۷۸	۱۱	۱۹۲	۱/۶۴	۲۵/۱۸	۱۲۲	۹۲	۱/۱۸	۳۶/۱۳	۲۰	۱۶۷	-۰/۳۳	۵۳/۲۸	۱۷۵	۶۷
۲/۲۷	۱۵/۶۹	۱۶۲	۱۹۳	۱/۶۶	۲۴/۸۱	۱۴۵	۹۳	۱/۱	۳۵/۷۶	۱۹۵	۱۶۸	-۰/۳۷	۵۲/۹۱	۱۸۰	۶۸
۲/۲۷	۱۵/۶۹	۱۵۶	۱۹۴	۱/۷۳	۲۳/۷۲	۱۷۷	۹۴	۱/۱	۳۵/۷۶	۱۶۰	۱۶۹	-۰/۳۷	۵۲/۱۸	۱۴۰	۶۹
۲/۳۶	۱۴/۵۹	۲۶	۱۹۵	۱/۷۷	۲۲/۹۹	۱۸۱	۹۵	۱/۱۲	۳۵/۴	۲۰۰	۱۷۰	-۰/۳۷	۵۲/۱۸	۱۲۹	۷۰
۲/۴۲	۱۳/۸۶	۱۸۳	۱۹۶	۱/۹۳	۲۰/۴۳	۱۸۷	۹۶	۱/۱۵	۳۴/۶۷	۱۰۹	۱۷۱	-۰/۳۷	۵۲/۱۸	۶۸	۷۱
۲/۵۵	۱۲/۴	۲۴	۱۹۷	۱/۹۸	۱۹/۷	۷۸	۹۷	۱/۱۷	۳۴/۳	۱۹۹	۱۷۲	-۰/۳۸	۵۱/۸۲	۱۹۴	۷۲
۲/۷۸	۱۰/۲۱	۱۷۳	۱۹۸	۲	۱۹/۳۴	۲۸	۹۸	۱/۱۷	۳۴/۳	۱۹۰	۱۷۳	-۰/۳۸	۵۱/۸۲	۱۳۸	۷۳
۲/۸۲	۹/۸۵	۷۷	۱۹۹	۲/۰۳	۱۸/۹۷	۳	۹۹	۱/۱۷	۳۴/۳	۱۷۹	۱۷۴	-۰/۳۸	۵۱/۸۲	۸۵	۷۴
۲/۶۳	۴/۷۴	۵۳	۲۰۰	۲/۱	۱۷/۸۸	۶۳	۱۰۰	۱/۱۹	۳۳/۹۳	۱۷۴	۱۷۵	-۰/۳۸	۵۱/۸۲	۸۰	۷۵

سوالاتی که از نظر پارامتر دشواری در یک سطح بودند، بروای یک محور قرار گرفتند. در این نمودار سوالات با پارامتر دشواری پایین در سمت راست و سوالات با پارامتر دشواری بالا در سمت چپ مرتب شدند. سوال ۵۳ آزمون با پارامتر دشواری (b=۳/۶۳) دشوارترین سوال شناسایی شد. از میان ۲۷۴ داوطلب شرکت کننده در آزمون تنها ۱۳ نفر معادل ۵ درصد داوطلبین به این سوال پاسخ درست داد و پارامتر توانایی معادل این دشواری برابر ۰/۷۸ بود. سوال ۱۰۲ با پارامتر دشواری (b=۳/۵۲) ساده‌ترین سوال در این آزمون بود. از میان ۲۷۴ داوطلب آزمون جامع پیش کارورزی، ۲۶۸ نفر معادل ۹۸ درصد داوطلبین به این سوال پاسخ درست دادند. توانایی برآورد شده براساس مدل راش برای این سوال برابر ۰/۴۸ بود.

همان‌گونه که در جدول شماره سه مشاهده می‌شود سوالات براساس پارامتر دشواری مدل راش مرتب شده‌اند. ساده‌ترین سوالات این آزمون به ترتیب سوال شماره ۱۰۲ با دشواری سوال ۷۲-۳/۵۲، سوال ۷۲ با دشواری سوال ۲/۸-۱۰۷ با دشواری سوال ۲/۷۱-۲/۷۸ بود. دشوارترین سوالات آزمون نیز سوال ۲/۸۲ با دشواری سوال ۲/۷۸، سوال ۷۷ با دشواری سوال ۱۷۳ با دشواری سوال ۳/۶۳ قرار داشت. جهت راهنمایی بیشتر داوران، دشواری سوالات براساس نظریه کلاسیک اندازه‌گیری نیز محاسبه و در اختیار آنان قرار گرفت. در دور دوم نقشه سوال آزمون در اختیار داوران قرار گرفت. نقشه سوال، نمودار هیستوگرامی از سوالات بود که توزیع سوالات آزمون بر حسب پارامتر دشواری نمایش داده شد.

(۷۰) قرار داشتند و میانگین پارامتر دشواری سوالات برابر (۴۶/۰=b) بود و نمره خام معادل پارامتر دشواری برابر نمره خام ۱۲۳ و پارامتر توانایی برابر (۵۸/۰) بود.

در مرحله‌ی سوم داوری، نمودار نقشه سوال در اختیار داوران قرار گرفت و آنان با اکثریت آراء روی ستون ۱۷ از نقشه سوال به توافق رسیدند. در این ستون سوالات (۸۶، ۵۷، ۴۴ و ...) بود.

جدول ۴: نظرات داوران آزمون پیش کارورزی نسبت به فرایند تعیین نمره برش

ردیف	عبارت	کاملاً موافق	کاملاً مخالف	کاملاً موافق	کاملاً مخالف
۱	درک روشنی از هدف برگزاری جلسه به دست آوردم.	۴۶/۱۵	۵۳/۸۵	۰/۰	۰/۰
۲	اطلاعات مربوط به نحوه تعیین استاندارد واضح و روشن بود.	۶۹/۲۳	۲۳/۰۸	۷/۶۹	۷/۶۹
۴	اطلاعات مربوط به تعیین استاندارد به روش نقشه سوال واضح و روشن بود.	۶۱/۵۴	۳۰/۷۷	۷/۶۹	۷/۶۹
۵	توصیف دانشجوی مرزی در تعیین استاندارد واضح و مفید بود.	۴۶/۱۵	۴۶/۱۵	۷/۶۹	۷/۶۹
۶	زمان کافی برای بحث درمورد «دانشجوی مرزی» در نظر گرفته شده بود.	۵۳/۸۵	۳۸/۴۶	۷/۶۹	۷/۶۹
۷	در بکاربردن روش نقشه سوال به خوبی عمل کردم.	۰/۰۰	۶۶/۶۷	۳۳/۳۳	۰/۰۰
۸	ارائه داده‌های واقعی در دستیابی به تعیین استاندارد برایم مفید بود.	۳۰/۷۷	۶۱/۵۴	۷/۶۹	۷/۶۹
۹	بحث بعد از اولین مرحله روش نقشه سوال برایم مفید بود.	۳۸/۴۶	۶۱/۵۴	۰/۰۰	۰/۰۰

اولین مرحله در نمره برش، تعیین داوطلب با حداقل صلاحیت (MCC) می‌باشد. وانگ (Wang) در پژوهش خود بیان داشت که با این که روش آنگوف رایج‌ترین روش تعیین نمره برش در آزمون‌های اعطای گواهی‌نامه است اما با توجه به مشکل شناختی این روش مخصوصاً در تعیین و درک داوطلب با حداقل صلاحیت از روش نقشه سوال استفاده کرده است. وی شواهد رویه‌ای برای استفاده از روش نقشه سوال ارائه می‌دهد و بیان می‌دارد در روش نقشه سوال داوران درک بهتری از داوطلب با حداقل صلاحیت کسب نموده‌اند و برآورد صحیحی از پارامتر دشواری سوال به دست آورده‌اند. استنابل (Schnabel) نیز با بررسی رویه‌ای روش نقشه سوال بیان می‌دارد داوران به خوبی مفهوم داوطلب با حداقل صلاحیت را درک کرده‌اند (۳، ۸). در پژوهش حاضر نیز داوران در طی جلسات توجیهی به درک روشنی از داوطلب با حداقل صلاحیت دست یافته‌اند.

از نظر وانگ و استنابل روش نقشه سوال، تکنیکی را ارائه می‌نماید که به کاهش پیچیدگی شناختی فرآیند تصمیم‌گیری داوران منجر می‌گردد و فرآیند تفکر داوران در تعیین احتمال پاسخ صحیح به سوالات را ساده می‌کند. ارائه پاسخ بله/ خیر به این سوال که: آیا (MCC) دارای توانایی در سطح ۵۰ درصد پاسخ صحیح به سوال است؟ در این روش فعالیتی ساده‌تر نسبت به روش‌های رایج دیگر است. بکارگیری نقشه سوال مبتنی بر مدل راش تصویر کلی از سوالات برروی هیستوگرام ارائه می‌نماید. نمره برش با شناسایی نقطه‌ای برروی هیستوگرام انجام می‌گردد که احتمال ۵۰ درصد پاسخ صحیح به سوال وجود دارد. در این روش مقدار احتمال ۵۰ درصد در

همان‌گونه که در جدول ۴ مشاهده می‌گردد، نتایج بررسی روایی، رویه‌ای روش تعیین استاندارد و نمره برش نشان داد که داوران در طی جلسه توجیهی روش نقشه سوال، درک روشنی از مفاهیم تعیین استاندارد، روش نقشه سوال و داوطلب با حداقل صلاحیت کسب نمودند. از نظر داوران، ارائه داده‌های واقعی توانسته به آنان در درک نمره برش کمک نماید ولی هنوز در کاربرد نقشه سوال به رضایت بالایی دست نیافتداند.

بحث و نتیجه‌گیری

دانشگاه‌های علوم پزشکی هر سال دو بار اقدام به برگزاری آزمون جامع پیش کارورزی می‌نمایند. نمره قبولی در این آزمون‌ها به روش هنجار مرجع تعیین می‌گردد، درصورتی که آزمون پیش کارورزی با هدف سنجش توانایی‌ها و شایستگی‌های حرفه‌ای دانشجویان رشته‌های پزشکی جهت ورود آنان به دوره اینترنتی اجرا می‌گردد و جزو دسته آزمون‌های ملاک مرجع می‌باشد. در اجرای این آزمون‌ها تعیین نمره برش از اهمیت خاصی برخوردار بوده و مجوز ورود دانشجویان به دوره اینترنتی می‌باشد. در این مطالعه پس از آموزش در زمینه مفهوم نمره برش و معرفی روش نقشه سوال، از داوران خواسته شد تا نمره برش را تعیین نمایند. پس از مرحله اول و دوم داوری آن‌ها به بحث و در مرحله سوم داوران تبادل نظر به یکدیگر پرداختند و در حین بحث با یکدیگر به تعیین نمره برش اقدام نمودند. توافق بین داوران بعد از بحث و بعد از بررسی پارامتر دشواری سوال و ارائه هیستوگرام پارامتر دشواری سوالات افزایش یافت. این امر، یافته‌های مطالعات دیگر را مبنی بر این که بحث و بررسی در راستای افزایش توافق بین داوران مفید هستند، تأیید می‌کند (۳۱، ۸، ۳).

است (۳۳). محدودیت دیگر مطالعه حاضر، داوران متخصص در رشته‌های پزشکی بود که در جلسه تعیین نمره برش در نقشه روانسنج به تعیین داوطلب با حداقل صلاحیت پرداختند و در جایگاه روانسنج قرار گرفتند که در این مطالعه سعی گردید با اجرای جلسه توجیهی و راهنمایی یک راهنمای در رشته سنجش و اندازه‌گیری ابهامات داوران کاوش یافته و فرآیند داوری تسهیل گردد. پیشنهاد می‌گردد در پژوهش‌های آینده به بررسی روایی و پایایی روش تعیین نمره برش به روش نقشه سوال و مقایسه آن با روش‌های متداول دیگر همانند آنگوف و بوک مارک پرداخته شود.

قدرتانی

این مقاله برگرفته از پایان‌نامه دوره دکتری دانشکده روانشناسی علوم تربیتی دانشگاه علامه طباطبائی تهران در ۱۳۹۵/۰۹/۰۱ رشته سنجش و اندازه‌گیری است که در جلسه شماره ۴۵۹ شورای پژوهشی دانشکده روانشناسی و علوم تربیتی دانشگاه علامه طباطبائی به تصویب رسیده است. هم‌چنین از دکتر سارا مرتاض هجری که در اجرای پنل داوران، همکاری داشتند کمال تشکر را دارد.

نقطه‌ای بر هیستوگرام قرار دارد که پارامتر توانایی داوطلب با پارامتر دشواری سوال برابر باشد (۸). در پژوهش حاضر نیز داوران از ارائه جدول پارامتر دشواری سوالات و همچنین تصمیم‌گیری براساس نقشه هیستوگرام رضایت داشتند.

بوکندا (Buckendahl) (Buckendahl) و همکارانش در پژوهش خود به مقایسه دو روش آنگوف و بوک مارک پرداختند و نتایج پژوهش آنان نشان داد که توافق داوران در دور دوم و سوم نسبت به دور اول بیشتر شده بود (۳۲). در پژوهش حاضر نیز در مرحله سوم هم گام با ارائه نقشه سوال توافق داوران نسبت به تعیین نقطه برش بیشتر از دو مرحله قبل بود. در دیگر مطالعات، محققان به این نتیجه دست یافتند که با استفاده از مدل راش، زمان تصمیم‌گیری کوتاه‌تر می‌گردد (۳۲، ۸). در پژوهش حاضر نیز داوران از زمان اجرای روش نقشه سوال رضایت داشتند.

یکی از محدودیت‌های این پژوهش اجرای روش نقشه سوال برای اولین بار در دانشگاه‌های علوم پزشکی کشور بود و داوران تجربه‌ایی در زمینه‌ی روش نقشه سوال نداشتند و در انجام این روش با ابهام روبرو بودند. در پژوهش‌های مربوط به تعیین نمره برش، روش آنگوف در آزمون‌های جامع دانشگاه‌های علوم پزشکی کشور مورد استفاده قرار می‌گرفته

References

- Epstein RM, Hundert EM. Defining and assessing professional competence. *Jama* 2002; 287(2): 226- 35.
- Kane MT. Examinee- centered vs. task-centered standard setting. In the Proceedings of the Joint Conference on Standard Setting for Large-scale Assessment 1995; 2, 119- 141.
- Schnabel SD. A Comparison of the Angoff and Item Mapping Standard Setting Methods for a Certification Examination [dissertation]. Illinois University: 2018.
- Cizek GJ, Setting performance standards: Foundations, methods, and innovations. Routledge; 2012.
- Çetin S, Gelbal S. A Comparison of Bookmark and Angoff Standard Setting Methods. *Educational Sciences: Theory and Practice* 2013; 13 (4): 2169- 75.
- Karantonis A, Sireci SG. The bookmark standard- setting method: A literature review. *Educational Measurement: Issues and Practice* 2006; 25 (1): 4- 12.
- Nasri Kh, Kahbazi M, Nasri Sh. [Medical Students' Viewpoints toward Basic Sciences and Preinternship Comprehensive Exams in Arak University of Medical Sciences]. *Iranian J Med Educ* 2010; 10 (1): 82- 91. [Persian]
- Wang N. Use of the Rasch IRT model in standard setting: An item- mapping method. *Journal of Educational Measurement*. 2003 Sep; 40 (3): 231- 53.
- Jaeger RM, Mullis IV, Bourque ML, Shakrani S. Setting performance standards for performance assessments: Some fundamental issues, current practice, and technical dilemmas. *Technical issues in large-scale performance assessment* 1996; 79- 115.

10. Jiao H, Lissitz RW, Macready G, Wang S, Liang S. Exploring levels of performance using the mixture Rasch model for standard setting1. *Psychological Test and Assessment Modeling* 2011; 53 (4): 499.
11. Yim M. Comparison of results between modified- Angoff and bookmark methods for estimating cut score of the Korean medical licensing examination. *Korean journal of medical education* 2018; 30 (4): 347.
12. Hambleton RK. Setting performance standards on educational assessments and criteria for evaluating the process. In *Setting performance standards*. Routledge; 2013.
13. Kaftandjieva F. Methods for setting cut scores in criterion-referenced achievement tests. Cito, Arnhem: EALTA; 2010.
14. Violato C, Marini A, Lee C. A validity study of expert judgment procedures for setting cutoff scores on high-stakes credentialing examinations using cluster analysis. *Evaluation & the health professions* 2003; 26 (1): 59- 72.
15. Bin Khairani AZ, bin Abd Razak N, Shamsuddin H. Application of the Rasch model and the bookmark method in setting cut scores in mathematics. *International Journal of Information and Education Technology* 2014; 4 (2): 198.
16. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Sage; 1991.
17. Näsström G, Nyström P. A comparison of two different methods for setting performance standards for a test with constructed- response items. *Practical Assessment, Research & Evaluation* 2008; 13 (9): 1- 12.
18. Cizek GJ, Bunch MB. Standard setting: A guide to establishing and evaluating performance standards on tests. SAGE Publications Ltd; 2007.
19. Beuk CH. A method for reaching a compromise between absolute and relative standards in examinations. *Journal of Educational Measurement* 1984 ; 21 (2): 147- 52.
20. Buckendahl CW, Smith RW, Impara JC, Plake BS. A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational measurement*: 39 (3): 253- 63.
21. Lee G, Lewis DM. A generalizability theory approach to standard error estimates for bookmark standard settings. *Educational and Psychological Measurement* 2008; 68 (4): 603- 20.
22. Cizek GJ. *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
23. Lypson ML, Downing SM, Gruppen LD, Yudkowsky R. Applying the Bookmark method to medical education: Standard setting for an aseptic technique station. *Medical teacher* 2013; 35 (7): 581- 5.
24. Plake BS. Setting performance standards for professional licensure and certification. *Applied Measurement in Education* 1998 1; 11 (1): 65- 80.
25. Chang L. Judgmental item analysis of the Nedelsky and Angoff standard-setting methods. *Applied Measurement in Education* 1999; 12 (2): 151- 65.
26. Pellegrino JW, Jones LR, Mitchell KJ. *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. National Academy Press; 1999.
27. Shepard LA. Implications for standard setting of the National Academy of Education evaluation of the National Assessment of Educational Progress achievement levels. InProceedings of the Joint Conference on Standard Setting for Large- Scale Assessments 1995; 2: 143- 160.

28. Abbasi H. [Comprehensive assessment and determination of the standards of mastery in specialized examinations for entry into internship courses in the medical field using the classic and featured model] [dissertation]. Allame Tabatabaei University; 2013. [Prsian]
29. Linacre JM. Winsteps Rasch measurement computer program user's guide. Beaverton, Oregon: Winsteps. Com; 2012.
30. Minaei A. [Application of Rash Measurement Model in Evaluation of Measurement Measurement Features of Visual- Motion Skills Test (TVMS- R)]. Educational Measurement Quarterly 2015; 5 (18): 77- 114.[Persian]
31. Yudkowsky R, Downing SM, Wirth S. Simpler standards for local performance examinations: the Yes/No Angoff and whole- test Ebel. Teach Learn Med 2008; 20 (3): 212- 7.
32. Peterson CH, Schulz EM, Engelhard Jr G. Reliability and validity of bookmark- based methods for standard setting: comparisons to Angoff- based methods in the National Assessment of Educational Progress. Educational Measurement: Issues and Practice 2011; 30 (2): 3- 14.
33. Mortaz Hejri S, Jalili M, Labaf A. [Setting Standard Threshold Scores for an Objective Structured Clinical Examination using Angoff Method and Assessing the Impact of Reality Chacking and Discussion on Actual Scores]. Iranian Journal of Medical Education. 2012; 11 (8): 885- 94.[Persian]

The Use of Rasch and Item Mapping in Determining Cut Score of Comprehensive Pre internship Exam

Sh Habibadeh¹, Delavar A^{2*}, Farrokhi NA³, Minaei A⁴, Jalili M⁵

Received: 2019/04/22

Accepted: 2019/07/22

Abstract

Introduction: One of the most controversial issues in educational measurement is to determine the cut score during the development and administration of a test. This paper aimed at using the Rasch model (item mapping) in determining cut score of the comprehensive pre internship exam in medical education.

Methods: This is a descriptive cross-sectional study. in doing so 13 faculty members of Tehran University of Medical Sciences were selected as the panelists Data collection instrument was the 200-item pre intership exam which administered in the Region 10 of the national medical science system between August 22 and September 21, 2016, that were responded by 274 students of university of medical sciences A cut score was set to identify pass/fail score for performance assessment. Data were analyzed by Winstep.

Result: According to the findings, the item reliability of this exam was 0.98 ($PR= 0.98$) and person reliability of 0.93 ($IP= 0.93$), that indicate the sample variance and test length were suitable and the items were selected from the nine regions proportionally. Examining items by panelists using the item map method showed three judgment stages included the item difficulty of (-0.46), raw score of 103, and ability parameter equivalent of (0.58).

Conclusion: Procedural validity of the cut score showed that the use of item map method resulted in greater agreement between the panelists on this score. Results indicated high satisfaction of the panelists because of feasibility of this process and implementation stag.

Keywords: Cut Score, Rosch Model, Item Mapping, Comprehensive Preinternship, Exam

Corresponding Author: Delavar A, Department of Assessment & Measurment, Faculty of psychology & Education, Allame tabatabaei University, Tehran, Iran delavarali@yahoo.com

Habibzadeh Sh, Ph.D Students of Assessment & Measurment, Faculty of psychology & Education, Allame tabatabaei University, Tehran,Iran

Farrokhi NA, Department of Assessment & Measurment, Faculty of psychology & Education, Allame tabatabaei University. Tehran, Iran

Minaei A, Department of Assessment & Measurment, Faculty of psychology & Education, Allame tabatabaei University, Tehran, Iran

Jalili M, Department of Emergency Medicine, Department of Medical Education, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran