

## تعیین حد نصاب قبولی راهکاری برای دستیابی به عدالت در ارزیابی: مطالعه‌ی مروری

عاطفه حیدری<sup>۱</sup>، شهرام شایان<sup>۲</sup>

## مقاله مروری

## چکیده

**مقدمه:** یکی از موضوعات چالش‌برانگیز در آموزش، برقراری عدالت در ارزیابی دانشجویان است. تعیین نمره‌ی قبولی ثابت برای همه‌ی آزمون‌ها، بدون در نظر گرفتن ویژگی‌های خاص هر آزمون و آزمون‌شونده، منصفانه نیست. برای برقراری عدالت، باید استاندارد قبولی هر آزمون بر اساس عواملی مانند سطح دشواری آزمون، توانایی دانشجویان، محیط یادگیری و اهداف آموزشی تعیین شود. این پژوهش با هدف ارائه‌ی راهکارهایی برای تعیین دقیق حدنصاب نمره‌ی قبولی و بهبود کیفیت آزمون‌ها انجام شده است. انتظار می‌رود نتایج این مطالعه بتواند به مدیران برگزارکننده‌ی آزمون‌ها و اعضای هیأت علمی دانشگاه‌ها کمک کند تا فرایند ارزیابی را بهبود بخشیده و در جهت برقراری عدالت آموزشی گامی مهم بردارند.

**روش‌ها:** این مطالعه از نوع مروری روایتی و جستجوی نظام‌مند بوده است. مقالات چاپ شده فارسی و انگلیسی با کلیدواژه‌های تعیین استاندارد (Standard setting)، نمره‌ی قبولی، حد نصاب قبولی (Minimum pass)، نقطه برش (Cut point) در پایگاه‌های اطلاعاتی نظیر PubMed, Google Scholar, Scopus, SID جستجو و در نهایت ۳۷ مطالعه بر اساس معیارهای ورود و خروج وارد مطالعه شدند.

**یافته‌ها:** مقالات وارد شده در مطالعه، ۱۶ روش تعیین حد نصاب قبولی را شرح داده بودند. به طور کلی این موارد در چهار دسته آزمون‌محور، آزمون‌شونده محور، کارایی محور و نتیجه محور تقسیم‌بندی می‌شوند.

**نتیجه‌گیری:** دانشگاه‌ها برای بهبود کیفیت آزمون‌های خود، با بهره‌گیری از متخصصین حوزه‌ی ارزیابی و امتحانات، اقدام به نظارت بر این فرایند نمایند و با استفاده از روش‌های ترکیبی و یا نقشی سؤال، آن را به عنوان یک رویکرد استاندارد در جهت دستیابی به عدالت در ارزیابی‌ها، مورد توجه قرار دهند.

**واژگان کلیدی:** اندازه‌گیری آموزشی؛ عدالت اجتماعی؛ آموزش پزشکی

**ارجاع:** حیدری عاطفه، شایان شهرام. تعیین حد نصاب قبولی راهکاری برای دستیابی به عدالت در ارزیابی: مطالعه‌ی مروری. مجله دانشکده پزشکی اصفهان ۱۴۰۴؛ ۴۳ (۸۰۲): ۵۳-۶۲.

## مقدمه

وجود حرفه‌ی پزشکی و علوم مرتبط با آن در هر جامعه‌ای ضروری است؛ بنابراین تربیت نیروی انسانی کارآمد و شایسته در آموزش پزشکی همواره مورد توجه بوده است. در دهه‌های اخیر استفاده از روش‌های ارزشیابی دانشجویان در علوم پزشکی پیشرفت‌های بسیاری داشته است و ما شاهد دستاوردهای علمی قابل قبول در این حوزه بوده‌ایم. یک ارزشیابی موفق باید اطمینان دهد که دانشجویان از دانش مناسب، مهارت، نگرش لازم برخوردار می‌باشند. لذا امتحان، ابزاری برای ارزشیابی دانشجویان و سنجش میزان مفاهیم فراگرفته شده است. هر چند روش‌های ارزشیابی متفاوت بوده و ممکن است به صورت

کتبی یا عملکردی برگزار شود اما در همه‌ی آن‌ها باید معیاری دقیقی جهت قضاوت شایستگی دانشجویان وجود داشته باشد. آنچه که همواره در برگزاری آزمون‌ها مورد توجه دانشجویان، اعضای هیأت علمی و مدیران برگزارکننده‌ی آزمون‌ها قرار دارد، رعایت عدالت در نمرده‌ی می‌باشد. امروزه آزمون‌ها نه تنها تأثیر مستقیمی در سرنوشت داوطلبان ایفا می‌کنند بلکه منجر بر بروز تأثیرات عمیق در لایه‌های مختلف اجتماعی می‌گردد. لذا توجه جدی به برگزاری مطلوب آزمون‌ها بخصوص آزمون‌های مهم کشوری از مطالبات جدی جامعه محسوب می‌شود (۱).

در این ارتباط تعیین نمره‌ی قبولی در آزمون، یکی از الزامات مهم

۱- دانشجوی کارشناسی ارشد، گروه آموزش پزشکی، مرکز مطالعات و توسعه آموزش پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.

۲- استادیار، گروه آموزش پزشکی، مرکز مطالعات و توسعه آموزش پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.

نویسنده‌ی مسؤول: شهرام شایان؛ استادیار، گروه آموزش پزشکی، مرکز مطالعات و توسعه آموزش پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

Email: shahram\_shayan10@yahoo.com

پزشکی می‌گذرانند. در پایان این مقطع، این امتحان به صورت دانش‌گاهی برگزار می‌گردد و به عنوان پیش‌نیاز ورود به مقطع فیزیوپاتولوژی است؛ همچنین از دیدگاه دانشجویان این آزمون جهت سنجش اطلاعات ضروری مورد استفاده قرار می‌گیرد (۹). بر اساس مطالعات نشان داده شده، جهت تعیین حد نصاب قبولی در این آزمون از روش خاصی استفاده می‌شود. به اینصورت که نقطه برش این آزمون، بستگی به نمرات پنج درصد افرادی است که بالاترین نمره را کسب می‌نمایند و معمولاً حداقل نصاب برابر با ۷۰ درصد میانگین نمرات پنج درصد فوق قرار داده می‌شود (۱۰).

علاوه بر این، پزشکان متخصص پس از اتمام دوره تخصص و یا فوق تخصص خود و برای عضویت به عنوان هیأت علمی دانشگاه‌های علوم پزشکی سراسر کشور ملزم به گذراندن آزمون دانشنامه‌ی تخصصی و فوق تخصصی پزشکی خواهند بود. این آزمون یکی از امتحانات مهم کشوری در جهت ارتقاء فرایند آموزش پزشکی است و در صورت اجرای استاندارد آن تأثیرات اجتماعی عمیقی به همراه خواهد داشت (۱۱). این آزمون ۱ یا ۲ بار در سال به صورت کتبی و شفاهی برگزار می‌گردد. اگر داوطلبان بتوانند حد نصاب لازم در آزمون کتبی و شفاهی را به دست بیاورند، مدرک دانشنامه یا بورد به آنها اعطا می‌شود. حد نصاب نمره قبولی جهت کسب مدرک دانشنامه‌ی تخصصی ۷۰ درصد کل نمره کتبی و ۷۰ درصد کل نمره شفاهی می‌باشد (۱۲).

علاوه بر آزمون‌های کتبی، برگزاری امتحانات عملکردی در حوزه علوم پزشکی نیز اهمیت فراوانی دارد. یکی از آزمون‌های عملی که به صورت معمول در رشته‌های علوم پزشکی جهت ارزیابی توانایی و مهارت دانشجویان استفاده می‌شود؛ آزمون بالینی ساختارمند عینی (OSCE) است. برخی از مطالعات نشان می‌دهد که استفاده از روش رگرسیون مرزی برای تعیین نقطه‌ی برش این آزمون‌ها، مناسب‌تر از سایر روش‌هاست و از پایایی و روایی بیشتری برخوردار است (۳). در آزمون‌های مرتبط با فارغ‌التحصیلی نیز، تعیین حد نصاب قبولی امری مهم است و 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 (۱).

در یکی از مقالات منتشر شده توسط خبرنگار شماره‌ی ۱۷ دبیرخانه شورای آموزش پزشکی و تخصصی در مورد تعیین حد نصاب آزمون، عنوان شد که با توجه به مشکلات اجرایی، انجام روش‌های علمی برای تعیین استاندارد عملاً امکان‌پذیر نیست و برای آزمون‌های حساس و مهم مانند ارتقا و دانشنامه هنوز هم از روش نمره ثابت استفاده می‌شود (۱۳).

پژوهش‌های متعددی در کشور طی سال‌های اخیر در مورد تعیین

و بحث‌برانگیز جهت حفظ ارزش و اعتبار آزمون محسوب می‌شود (۱). تعیین حداقل نمره‌ی قبولی (نقطه‌ی برش یا تعیین استاندارد) عبارت است از روند نظام‌بندی که طی آن مشخص می‌گردد، دانشجویان باید چه نمره‌ای کسب نمایند تا در آزمون قبول شوند (۲). در واقع نمره‌ای است که آزمون‌دهندگان را به طبقات مختلف (نظیر رد- قبول) تقسیم می‌کند. اکثراً نقطه‌ی برش در آزمون‌های پیشرفت تحصیلی در مقطع کارشناسی، ۱۲ و کارشناسی ارشد و دکتری برای دروس اختصاصی، ۱۴ می‌باشد (نمره‌ی ثابت) و پایین‌تر از این نمره به منزله‌ی مردودی است (۳، ۴).

با توجه به اینکه در آزمون‌های مختلف شرایط، میزان دشواری سؤالات، خصوصیات و سطح فراگیران باهم متفاوت است؛ بنابراین عادلانه و منطقی نیست که حد نصاب قبولی در همه‌ی آزمون‌ها یکسان باشد و بایستی حد نصاب قبولی برای همان آزمون تعیین و با دقت و حساسیت انجام شود تا قضاوت‌هایی که صورت می‌گیرد، کمترین خطا و لطمه به دانشجویان را وارد کند (۵). از آنجا که تعیین نقطه برش یا حد نصاب قبولی برای تمایز قابل شدن بین دانشجویان توانمند از غیر توانمند است، بنابراین تعیین آن باعث خواهد شد افرادی که از قابلیت و توانمندی بیشتری برای انجام وظایف محوله در فرایند درمان بیماران و نیز حفظ و ارتقای سلامت آحاد جامعه برخوردار هستند، اجازه ورود به عرصه را دریافت نمایند. خطا در زمینه‌ی تعیین و طبقه‌بندی افراد با حداقل شایستگی می‌تواند هزینه‌های زیادی را به همراه داشته باشد. از طرفی سختگیری زیاد در این زمینه، باعث عدم موفقیت افراد واجد شرایط می‌شود و از سوی دیگر سهل‌گیری بی‌جا، هزینه‌های اجتماعی زیادی را به همراه خواهد داشت (۶). در این راستا روش‌های مختلفی برای تعیین نقطه‌ی برش تعیین شده است اما هیچ کدام از آنها روش کاملاً بی‌نقصی برای تعیین نمره‌ی برش محسوب نمی‌شوند و نمی‌توان یکی از آنان را به عنوان بهترین روش معرفی کرد؛ بلکه بسته به مزایا و معایب خاص هر کدام، در موقعیت‌های مختلف به کار برده می‌شوند (۷). همچنین عوامل مختلفی در تعیین نمره‌ی برش تأثیر می‌گذارند به عنوان مثال مطالعات نشان داده که حتی تغییر در قالب سؤال در تعیین نقطه‌ی برش تأثیر گذاشته است (۸)؛ بنابراین انتخاب روش مناسب امری ضروری است. که این مهم نیازمند بهره‌گیری از متخصصین حوزه مربوطه می‌باشد.

دانشجویان رشته پزشکی در طول دوره‌ی هفت ساله خود، سه امتحان علوم پایه، فیزیوپاتولوژی و صلاحیت بالینی را برای دریافت مدرک پزشکی عمومی پشت سر می‌گذارند تا بتوانند در آزمون دستیاری (تخصص) شرکت کنند. آزمون علوم پایه یکی از مهم‌ترین ارزشیابی‌های موجود در حوزه علوم پزشکی است (دانشجویان رشته‌ی پزشکی در مقطع علوم پایه، حدود ۶۸ واحد درسی پایه

### طبقه‌بندی روش‌های اندازه‌گیری حد نصاب قبولی

به طور کلی آزمون‌ها بر اساس شیوهی نمره‌دهی به دو دسته کلی تقسیم می‌شوند: ملاک مرجع و هنجار مرجع.

- **آزمون‌های هنجار مرجع:** ملاک قبولی و مردودی دانشجویان بعد از آزمون مشخص می‌شود و عملکرد کل دانشجویان را در نظر می‌گیرد. مثلاً فردی که نمره او ۶۰ در صد بالای گروه باشد قبول می‌شود. این روش بیشتر برای انتخاب تعداد محدودی از آزمون‌شوندگان، مانند پذیرش در کنکور استفاده می‌شود.

- **آزمون‌های ملاک مرجع:** آزمون‌هایی که بیشتر بر اهداف یادگیری تأکید می‌کند و اینکه آیا آزمون‌شونده در آزمون موفق می‌شود یا مردود می‌شود، به سطح دانش و مهارت‌های او در مقایسه با معیارهای کاملاً تعریف شده بستگی دارد. این نوع از نمره‌دهی بیشتر در امتحانات نهایی یا امتحانات فارغ التحصیلی مورد استفاده قرار می‌گیرد، جایی که «میزان دانش یا مهارتی که یک آزمون‌شونده واقعاً می‌داند یا می‌تواند انجام دهد» حیاتی است (۱۵، ۱۶). به عنوان مثال از این روش، در یادگیری در حد تسلط بسیار استفاده شده و تعیین نقطه‌ی برش برای تعیین اینکه چه زمانی یادگیرنده به تسلط کافی رسیده، حیاتی است (۱۷، ۱۸).

ترنیول عقیده داشت که «از آنجایی که هدف اصلی آموزش پزشکی تولید یک پزشک شایسته است، پس بی‌تردید دلیل وجودی فرایند ارزشیابی، ارزیابی شایستگی و نه رتبه‌بندی دانشجویان است» (۱۵، ۱۹). بنابراین، استفاده از آزمون‌های ملاک مرجع برای تعیین سطح مورد نیاز از شایستگی مناسب‌تر خواهد بود (۱۵). علیرغم پیشرفت قابل توجهی که در زمینه تعیین نمره به شیوهی ملاک مرجع وجود دارد، بسیاری از اساتید پزشکی هنوز از روش‌های هنجار مرجع و تفکر هنجاری در آزمون‌ها استفاده می‌کنند (۱۹). یکی از مدرن‌ترین و قابل قبول‌ترین طرح طبقه‌بندی ارائه شده، طرح پیشنهادی جاگر است که روش اندازه‌گیری نقطه برش را با توجه به شاخصه تمرکز بر موضوع توضیح داده است و آن را به چهار دسته تقسیم کرد:

#### ۱. Test-centered: در این روش موضوع مورد قضاوت،

سؤال‌های آزمون می‌باشد (آزمون‌محور). در اکثر موارد جهت تعیین نقطه‌ی برش (حدود ۶۸ درصد) از این روش استفاده می‌کنند. از مزایای آن می‌توان به عدم نیاز به مهارت و آمادگی، زمان و منابع زیاد را نام برد. غالباً از این روش، قبل از اجرا آزمون استفاده شده و داوران زیادی می‌توانند در مورد سؤالات نظر دهند. این ویژگی سبب دستیابی به سطوح بالاتر پایایی در قضاوت می‌شود. در واقع نیاز به داور جهت داوری درباره سختی آزمون‌های آزمون دارد که با قضاوت در مورد احتمال دادن پاسخ

استاندارد قبولی انجام شده است. لیکن با توجه به جستجو در پایگاه‌های اطلاعاتی مختلف، این مطالعات به صورت پراکنده و دسترسی به آنها محدود بوده و در مورد آنها یک جمع‌بندی علمی همراه با ذکر نمونه‌های مختلف و ارائه‌ی راهکار جهت برقراری عدالت در ارزیابی صورت نگرفته است. در این مقاله به بررسی انواع روش‌های اندازه‌گیری حد نصاب قبولی، و ارائه‌ی راهکار جهت برقراری عدالت در ارزیابی دانشجویان به عنوان راهنمای مدیران آزمون و طراحان سؤال پرداخته شده و در تهیه‌ی آن بیش از ۱۰۰ مقاله انگلیسی و فارسی بررسی و از بین آنها ۳۷ مطالعه استفاده شده که در بخش منابع، مشخصات آن‌ها ارائه شده است.

### روش‌ها

مطالعه‌ی حاضر از نوع مروری روایتی و جست‌وجوی نظام‌مند بود. در این مطالعه، مروری جامع و عمیق با استفاده از پایگاه‌های داده علمی انجام شد. این مطالعه با هدف تعیین حد نصاب قبولی و شیوه‌های محاسبه آن انجام شده و یافته‌ها از طریق جستجو در پایگاه‌های اطلاعاتی انگلیسی و فارسی زبان PubMed, Google Scholar, Scopus SID با کلیدواژه‌های تعیین استاندارد (Standard setting)، حد نصاب قبولی (Minimum pass level)، نمره‌ی قبولی (Minimum pass)، نقطه برش (Cut point) استخراج شد. مطالعات از ابتدا تا تاریخ ۱۷ بهمن ۱۴۰۲ جمع‌آوری شدند. مقالات پژوهشی به صورت هدفمند بر اساس معیارهای ورود به مطالعه‌ی حاضر که شامل: مرتبط با هدف پژوهش، مرتبط با علوم پزشکی، در دسترس بودن متن کامل مقاله وارد مطالعه شدند.

در جستجوی اولیه بر اساس کلیدواژه‌ها بیش از ۱۰۰ مطالعه جمع‌آوری شد. با بررسی چکیده و متن مقالات توسط دو نفر از اعضای تیم پژوهش بر اساس معیارهای ورود و پس از حذف موارد تکراری و غیرمرتبط، ۳۷ مطالعه انتخاب شدند.

### یافته‌ها

#### تاریخچه

در سال ۱۹۶۶، دانشکده‌ی پزشکی دانشگاه میسوری دانشجویان را بر اساس عملکرد دانشجویان هر دوره به سه دسته‌ی کلی پاس/مردود/شاگرد اول تقسیم می‌کرد. گروه آسیب‌شناسی در دوره دوم، استفاده از این شیوه را نامناسب دیدند آن‌ها متوجه شدند که تنها با مشخص کردن اهداف نمی‌توانند در مورد قبولی یا مردودی قضاوت کنند زیرا اهداف به اندازه کافی دقیق نبودند؛ به این ترتیب استفاده از تکنیک تعیین حداقل نمره‌ی قبولی برای امتحانات چهارگزینه‌ای آغاز شد (۱۴).

دانشجوی مرزی بتواند به این سؤال پاسخ صحیح بدهد) جواب می دهد. پس از ارائه نظرات داوران میانگین پاسخ آنان محاسبه شده و به این صورت نقطه برش مشخص می شود. برای تعیین کل نمره‌ی آزمون، دوباره از میان تمام نقطه برش‌های سؤالات، میانگین گرفته می شود. همانطور که قبلاً اشاره شد با استفاده از روش انگوف می توان استناداری روا و پایا برای آزمون‌های بالینی ساختارمند عینی (OSCE) به دست آورد (۱، ۲۲)؛ و از آنجایی که ارزیابی واقعی و منصفانه صلاحیت‌های بالینی دانشجویان کاری سخت و دشوار است (۲۳) انتخاب معیاری مناسب جهت قضاوت صحیح از اهمیت بسیار بالایی برخوردار است.

### انواع انگوف

**انگوف تغییر یافته:** تفاوت آن با انگوف اصلی این است که امکان بحث میان داوران در فواصل کاری و در انتهای جلسه وجود دارد و داوران می توانند نمرات محاسبه شده را نقد یا حتی تغییراتی در آن اعمال کنند. در نوع دیگری از روش انگوف، پیشنهاد شده که قبل از شروع بحث نهایی، نمرات واقعی دانشجویان در اختیار داوران قرار گیرد. چون در این روش داوران با واقعیت روبه رو می شوند، راحت تر می توانند نقطه‌ی برش را تعیین نمایند (۱، ۲۴).

**انگوف بله/خیر:** در روش انگوف بله/خیر احتمال قبولی به صورت بله/خیر بیان می شود (۵).

**انگوف سه سطحی:** استفاده از سه پاسخ «بله»، «خیر» و «شاید» برای تعیین احتمال قبولی (۵).

**انگوف مبتنی بر نظریه‌ی سؤال پاسخ:** در این روش با استفاده از نظریه‌ی سؤال پاسخ (IRT) پارامترهای دشواری سؤالات تعیین و در کنار هر یک از سؤالات ضرایب دشواری (P value) آنها درج می شود تا داوران موجود در حوزه‌ی مربوطه احتمال پاسخ یک داوطلب مرزی به هر سؤال را محاسبه کنند؛ سپس از میان نمرات ارائه شده توسط ارزیابان میانگین گرفته خواهد شد. معمولاً این روش در سه مرحله اجرا می شود و در هر مرحله بازخوردها توسط داوران ارائه شده و در آخر نمره برش نهایی تعیین می شود (۵).

**بوکمارک:** یک رویکرد نسبتاً جدید است که در سال ۱۹۹۶ توسط یک گروه تحقیقاتی برای استفاده در آموزش دوره ابتدایی تا دبیرستان (k-12) در کانادا پیشنهاد شده است. در واقع از روش‌های سؤال‌محور برای تعیین حد نصاب قبولی است که در آن سؤالات به ترتیب ضریب دشواری براساس نظریه سؤال پاسخ نوشته می شوند و سپس لیستی طبقه بندی شده از سؤالات آسان به دشوار تهیه و در اختیار داوران قرار داده می شود (۲). سپس داوران در چندین مرحله احتمال پاسخ دانشجو دارای حداقل شایستگی را به هر یک از سؤالات مشخص می کنند. در نهایت میانگین احتمالات داده شده به هر یک از

صحیح در دانشجویان تعریف شده یا بوسیله‌ی طبقه‌بندی سوال های آزمون به چندین گروه از پیش تعیین شده بر حسب سطح شایستگی مورد نیاز تعیین می شود.

**۲. Examinee-centered:** این روش یکی از اولین روش های اندازه‌گیری حد نصاب بوده و معیار مورد نظر آن ویژگی های آزمون‌شوندگان می باشد. در اینجا برای هر آزمون‌شونده یک داور جهت شناسایی سطح شایستگی در اختیار است و کل رفتارهای دانشجو قبل از آزمون به عنوان شاخص مد نظر قرار می گیرد. این روش درجه کفایت و پایایی بالایی دارد؛ اما از نظر عملی کاربرد آن دشوار است و نمی تواند شایستگی دانشجویان را به طور کامل مورد سنجش قرار دهد.

**۳. Performance-centered:** معیار مورد نظر برای قضاوت، پاسخ دانشجویان به سؤال‌های آزمون به طور مجزا است (کارایی محور). در این روش، محصول فعالیت‌های آزمون‌شونده در طی آزمون مورد سنجش قرار می گیرد و دیگر نیازی به تعداد بالای داور ندارد.

**۴. Result-centered:** شاخص مورد بررسی در اینجا نمرات آزمون یا توزیع فراوانی آن‌هاست (نتیجه محور). از این روش برای آزمون‌هایی که چند قسمت دارند و مهارت‌های پیچیده اندازه‌گیری می شود، مورد استفاده قرار می گیرد (۱۳).

علاوه بر تقسیم‌بندی جاگر، تعیین حد نصاب نمره‌ی قبولی بر اساس معیار قضاوت به سه دسته دیگر نیز تقسیم می شود. دسته اول، قضاوت مبتنی بر آزمون‌دهندگان لب مرزی همانند روش ندلسکی، انگوف و روش بله خیر، دسته دوم قضاوت بر اساس ویژگی‌های هر یک از داوطلبان آزمون مثل روش گروه مرزی و گروه متضاد و دسته سوم سایر روش‌ها که هر کدام رویکرد متفاوتی برای تعیین نمره قبولی دارند (۲۰). صرف نظر از انواع دسته‌بندی‌هایی که برای روش‌های اندازه‌گیری حد نصاب نمره قبولی وجود دارد، در ادامه به روش‌هایی که جهت تعیین آن برای آزمون‌های مختلف به کار برده شده پرداخته می شود:

**انگوف:** یکی از مشهورترین و پرکاربردترین روش‌های تنظیم استاندارد روش انگوف می باشد. در این روش داوران به جای ارزیابی دانشجو، در مورد سؤالات آزمون نظر می دهند. متخصصان مسلط بر محتوای آزمون در مورد حداقل توانایی قابل قبول برای آزمودنی تصمیم‌گیری می کنند. این روش معیار منصفانه‌ای برای قضاوت محسوب می شود (۱۳، ۲۱).

در روش انگوف، ابتدا سؤالات طراحی شده و پانلی از متخصصان به عنوان داور (قبل از آزمون یا بعد از آن) مشخص می شود. هر داور به طور جداگانه به سؤال (چقدر احتمال دارد یک

سؤالات با همدیگر جمع شده و نمره‌ی برش برای آزمون حاصل می‌شود. روش بوکمارک امر قضاوت را با کاهش و یا تقویت بار شناختی ساده می‌کند (۲۲، ۲۵).

سؤالات با همدیگر جمع شده و نمره‌ی برش برای آزمون حاصل می‌شود. روش بوکمارک امر قضاوت را با کاهش و یا تقویت بار شناختی ساده می‌کند (۲۲، ۲۵).

**هافستی:** روش توافقی هافستی در سال ۱۹۸۳ به عنوان رویکردی نوآورانه پیشنهاد شد. این روش یکی از رویکردهای توافقی تعیین استاندارد آزمون است که در آن هم درصد مردودی و هم درصد قبولی افراد مرزی در آزمون مورد قضاوت قرار می‌گیرد. در حقیقت نتیجه‌ی نهایی، محصول مصالحه بین این دو قضاوت است و یک روش نتیجه‌محور می‌باشد. (۲۲).

به طور مثال در یک مطالعه برای محاسبه‌ی حد نصاب قبولی آزمون بالینی ساختارمند عینی از روش رگرسیون مرزی استفاده کرد که برای هر ایستگاه یک معادله رگرسیون طراحی شد. به این ترتیب که نمرات چک لیست به عنوان متغیر وابسته و نمرات کلی آزمون به عنوان متغیر مستقل در نظر گرفته شد؛ سپس با قرار دادن نمره‌ی کل مربوط به دانشجوی مرزی در معادله، حد نصاب قبولی برآورد شد (۳).

در این روش پس از تعیین اهداف آموزشی و آموزش دانشجویان، فراگیران از انتظارات داوران آگاهی می‌یابند. سپس چهار سؤال طراحی شده و از آن‌ها آزمون گرفته می‌شود. از میان این چهار سؤال، دو سؤال در سطح قابل قبول دانش است که با نماد K نشان داده می‌شود و دو سؤال دیگر بر میزان مردودی یا شکست در آزمون تمرکز دارد که با نماد F نشان داده می‌شود. چهار سؤال پرسیده شده از شرکت کنندگان به قرار زیر است:

**روش کوهن:** یکی از ساده‌ترین روش‌هاست که درصد مشخصی از نمرات بالای آزمون را به عنوان ملاک در نظر می‌گیرد (مثلاً ۶۰ درصد). در این روش چون نمره دانشجویان ممتاز ملاک می‌باشد؛ تعیین استاندارد به سطح توانایی شرکت‌کنندگان بستگی دارد (۲۸).

- بالاترین درصد قبولی که با نماد Kmax نشان داده می‌شود.  
- پایین‌ترین درصد قبولی که با نماد Kmin نشان داده می‌شود.  
- بالاترین درصد مردودی که با نماد Fmax نشان داده می‌شود.  
- پایین‌ترین درصد مردودی که با نماد Fmin نشان داده می‌شود.  
با محاسبه‌ی میانگین نظرات در هر یک از چهار مورد یاد شده، داده‌های حاصل به صورت نموداری نمایش داده می‌شود. نقاط حاصل از این نمودار، یک چهار ضلعی را تشکیل خواهند داد. وقتی قطر این چهار ضلعی ترسیم شد، استاندارد آزمون نقطه‌ای است که از تقاطع این قطر با منحنی عملکرد شرکت‌کنندگان در آزمون بدست می‌آید.

**روش گروه متضاد:** در این روش داوطلبان آزمون طی انجام خود ارزیابی، خود را بر اساس میزان مهارت و دانش در دو دسته واجد شرایط و غیر واجد شرایط تقسیم‌بندی می‌کنند. پس از تصحیح آزمون و توزیع نمرات بین آزمون‌دهندگان، نمودار توزیع نرمال نمرات برای دو گروه واجد شرایط و غیر واجد شرایط رسم می‌شود. نقطه تلاقی این دو نمودار، نقطه برش آزمون می‌باشد (۲۹).

برای اولین بار در الیپاد علمی دانشجویان علوم پزشکی از روش هافستی استفاده شده است. به دلیل اینکه این رویکرد سبب شفاف‌سازی و کاهش استرس شرکت‌کنندگان در آزمون می‌شود برای آزمون‌های رقابتی کشوری از اهمیت بالایی برخوردار می‌باشد (۲۶).

**روش ایبل (Ebel):** این روش برای سؤالات چهارگزینه استفاده می‌شود. همانند روش آنگوف کار برآورد تعداد آزمودنی‌ها با حداقل توانایی که می‌توانند یک سؤال را درست پاسخ دهند بر عهده اعضای هیأت داوران است. در روش Ebel، پانل متخصصان تشکیل شده؛ سپس هر سؤال را بررسی می‌کنند تا سطح دشواری (آسان، مناسب و دشوار) و مرتبط بودن (ضروری، مهم، قابل قبول و سؤال برانگیز) هر مورد را تعیین کند. سپس، هر آیتم با توجه به سطح دشواری و ارتباط در یک جدول ماتریس ۴×۳ طبقه‌بندی می‌شود. در مرحله بعد، داور درصد مورد انتظار به موارد موجود در هر خانه از جدول ماتریس را با توجه به داوطلب با حداقل شایستگی تعیین می‌کند. در نهایت، تعداد موارد در هر دسته در صد مورد انتظار پاسخ‌های صحیح ضرب می‌شود و کل نتایج برای محاسبه امتیاز برش جمع خواهد شد (۲، ۲۲، ۳۰). این روش با توجه به اینکه سطح دشواری و ارتباط محتوای سؤالات را مدنظر قرار می‌دهد، از ارزش بالایی برخوردار است. با این حال، در آزمون‌های با حجم بالای سؤال، تعیین نمره‌ی قبولی به چالشی جدی تبدیل می‌شود (۲۰).

**نمره‌ی ثابت:** یکی از معمول‌ترین و ساده‌ترین روش، در نظر گرفتن یک نمره‌ی ثابت برای آزمون است؛ همانطور که قبلاً هم اشاره شد نقطه برش در اکثر آزمون‌های پیشرفت تحصیلی در مقطع کارشناسی ۱۲، کارشناسی ارشد ۱۴ و دکتری ۱۶ می‌باشد. از معایب این روش آن است که به شرایط، نوع و سختی آزمون بی‌توجه است (۲۷).

**نیدلسکی:** روشی مورد استفاده برای سؤالات چهارگزینه‌ای است. در روش نیدلسکی ابتدا پانلی از متخصصان تشکیل می‌شود و از آن‌ها

**روش رگرسیون مرزی:** در این روش با استفاده از نمرات داوطلبان در آزمون، نمره‌ی حد نصاب قبولی تعیین می‌شود. در واقع، رگرسیون مرزی روشی است که به طور مستقیم بر عملکرد داوطلبان متمرکز است و به همین دلیل، به طور گسترده در آزمون‌های بالینی ساختارمند عینی OSCE کاربرد دارد. در این روش، به هر داوطلب بر

از آنجا که روش تعیین نمره قبولی بر دقت تشخیص این گروه مرزی تأثیرگذار است، لذا ضرورت توجیه و اثبات دقت ابزارهای سنجش مورد استفاده، بیش از پیش احساس می‌شود. به بیان دیگر مثلاً در امتحانات ۲۰ نمره‌ای دانشجویانی که نمره‌ی ۹/۷۵ یا ۱۰/۲۵ می‌گیرند تقریباً از نظر عملکرد همانند هم هستند؛ اما اگر امتحان مجدداً برگزار شود از کجا بدانیم که جای این دو باهم عوض نمی‌شوند؟ در واقع حد نصاب قبولی که داوران تعیین می‌کنند باید معنی‌دار باشد (۳۴).

در مطالعه‌ای برای بررسی معنی‌داری قبولی و مردودی دانشجویان، از روش کوکران که یک روش آماری است؛ استفاده کرد و دانشجویان را در سه طبقه قرار داد: قبولی معنی‌دار، مردودی معنی‌دار و غیر قابل قضاوت یا مشروط. در ست است که این روش معنی‌داری را مشخص می‌کند اما نیاز به تعداد بالای سؤال دارد. برای رفع محدودیت نیاز به تعداد بالای سؤال در روش کوکران، از محاسبه خطای آلفا به عنوان معیاری برای تعیین سطح اطمینان استفاده شد. همچنین، برای تصمیم‌گیری در مورد وضعیت دانشجویان مشروط، از روش‌های کیفی مانند مصاحبه بهره گرفته شد (۳۴).

#### تعیین دقت نمره‌ی برش

از آنجایی که تعیین نقطه‌ی برش در روش‌های مختلف متفاوت است؛ نیاز به شاخصی جهت تعیین دقت اندازه‌گیری داریم. معمولاً برای گزارش دقت اندازه‌گیری در مسائل مختلف از ضریب آلفای کرونباخ استفاده می‌شود اما این روش برای تصمیم‌گیری‌های دوگانه مانند قبولی و مردودی مناسب نمی‌باشد. از این رو بهتر است از خطای استاندارد شرطی و تعیین پایایی شرطی (Conditional reliability and conditional standard error of measurement (cSEM) استفاده شود (۳۳۵).

#### بحث

مطالعات متعددی درباره‌ی تعیین حد نصاب نمره‌ی قبولی انجام شده، اما این مطالعات به صورت پراکنده بوده و دسترسی به آنها محدود است. این پژوهش با بررسی مقالات متعدد، انواع روش‌های تعیین حد نصاب قبولی را بررسی کرده و به ارائه‌ی یک جمع‌بندی جامع از این مطالعات پرداخته است. در نهایت پیشنهادهایی به مسئولین برگزاری آزمون و اعضای هیأت علمی دانشگاه‌ها ارائه شده است.

ارزیابی دانشجوی یکی از ارکان مهم برنامه‌ریزی درسی است و به صورت مداوم در کنار آموزش انجام می‌شود (۳۶). نتایج این ارزیابی می‌تواند به اساتید و مسئولین جهت تصمیم‌گیری‌های مهم در مورد ابعاد مختلف فرایند یاددهی-یادگیری کمک نماید. از آنجا که فرایند ارزیابی دانشجوی حساسیت و اهمیت بالایی در جهت تصمیم‌گیری‌های

درخواست شده تا به بررسی هر یک از سؤالات بپردازند. در این روش داوران باید مشخص کنند که یک داوطلب سطح پایین چقدر احتمال دارد آن گزینه را به عنوان گزینه‌ی غلط تشخیص داده و سپس آن را حذف کند. از این طریق شانس پاسخگویی صحیح سؤال نیز مشخص می‌شود مثلاً اگر دو گزینه را بتواند حذف کند شانس پاسخگویی صحیح به سؤال ۵۰ درصد می‌باشد. در واقع در این رویکرد داوران به تشخیص عوامل خطا می‌پردازند. سپس نظرات هر یک از داوران بیان شده و در دور دوم نظر نهایی تصویب می‌شود (۳۱). این روش با وجود اینکه می‌تواند در آزمون‌های چند حوزه‌ای مورد استفاده قرار گیرد؛ نیازمند دانش تخصصی است و استفاده از آن برای همه‌ی آزمون‌ها مناسب نمی‌باشد (۲۰).

#### روش نقشه‌ی سؤال (Item Mapping) یا راش: مدل نقشه

سؤال برگرفته از روش بوکمارک می‌باشد. این روش از نظریه سؤال پاسخ بر اساس مدل راش برای فرایند تعیین استاندارد و نمره‌ی برش استفاده می‌نماید. با استفاده از مدل راش، داوران به این سؤال پاسخ می‌دهند که «آیا یک آزمون‌شونده با صلاحیت حداقلی به این آزمون پاسخ صحیح می‌دهد؟». تفاوت آن با روش آنگوف این است که برخلاف آنگوف که داوران به بررسی عملکرد داوطلبان در تک تک سؤالات می‌پردازند اما در روش نقشه سؤال یک تصویر کلی از همه‌ی سؤالات بر اساس پارامتر دشواری آورد شده را در هیستوگرام ارائه می‌نماید (۲).

#### روش EZ یا آسان: در این روش هر یک از اعضای پانل دور هم

جمع می‌شوند و شروع به تصحیح برگه‌های امتحانی می‌کنند؛ سپس دو نمره حداقل نمره‌ی قبولی و بالاترین نمره‌ی مردودی را مشخص کرده و به روش ZScore آن را تعیین می‌کنند. به دلیل عملیاتی و معتبر بودن این شیوه، به روش آسان نیز معروف است (۳۲).

#### روش جدید استفاده از معدل کل دانشجویان: پژوهش Aung

و همکاران روشی کارآمد برای تعیین حد نصاب قبولی ارائه کرده است که نیاز به حضور متخصصان موضوعی را مرتفع می‌سازد. در این روش با استفاده از معدل کل سال گذشته، دانشجویان مرزی را مشخص کرده و سپس میانگین نمره‌ی خام آن‌ها را در آزمون به عنوان نقطه برش آزمون در نظر می‌گیرد. در این پژوهش نمره‌ی حاصل از روش جدید با نمره‌ی برش روش آنگوف مقایسه شد. نتایج نشان داد که نمره‌ی قبولی تعیین شده توسط روش جدید با روش آنگوف همخوانی دارد و در مواردی که دسترسی به متخصص موضوعی نداریم؛ جایگزین مناسبی است (۳۳).

#### تعیین معناداری قبولی و مردودی: شایان ذکر است که تعیین

نمره قبولی توسط داوران، فارغ از روش مورد استفاده، منجر به ایجاد گروهی از دانشجویان می‌شود که با اختلاف ناچیزی در دو سوی مرز قبولی قرار می‌گیرند. عملکرد این دو گروه عملاً قابل تفکیک نیست.

بوکماریک، در روش ایبل داوران سطح دشواری و مرتبط بودن هر سؤال را خودشان تعیین می‌کنند. در روش‌هایی همانند هافستی، ندلسکی و کوهن خود دانشجویان به جای سؤالات آزمون مورد بررسی قرار می‌گیرد.

از طرفی انتخاب روش مناسب جهت تعیین نقطه‌ی برش به معیارهای مختلفی مثل سؤال‌های آزمون، ویژگی دانشجویان، پاسخ دانشجویان به سؤالات و نمرات آزمون بستگی دارد. علاوه بر این مطالعات نشان داده که خود نمره‌ی برش نیز تحت تأثیر عوامل مختلفی مثل ویژگی‌های ابزار اندازه‌گیری، داوران شرکت‌کننده جهت قضاوت و معیارهای عملکرد تعیین شده بستگی دارد. برای مقایسه دقیق این روش‌ها نیاز است تأثیر این عوامل را نادیده بگیریم یعنی برای یک آزمون با ثابت بودن داوران، معیارهای عملکرد، و ابزار اندازه‌گیری نمره‌ی برش محاسبه شود (۳۸).

پژوهش‌های مختلف نشان دادند که حد نصاب قبولی به دست آمده در روش‌های مختلف متفاوت است و هیچ روشی بدون عیب نیست و روشی که از همه لحاظ مورد قبول باشد در دسترس نمی‌باشد. بنابراین انتخاب بهترین روش کار دشواری است و برای تعیین آن باید به پایایی و روایی آن‌ها، ضرورت نیاز به قضاوت، شایستگی داوران درگیر در تعیین نقطه برش و غیره توجه کرد (۱)، (۱۳). با توجه به اهمیت زیاد تعیین نقطه برش به عنوان معیاری برای قضاوت دانشجویان در امتحانات پیشرفت تحصیلی و هم در آزمون‌های ورودی، مهم است که دست‌اندرکاران برگزار آزمون و داوران در زمینه‌ی اصول و فنون این حیطة تسلط کافی داشته باشند.

### نتیجه‌گیری

این مطالعه با تعیین حد نصاب قبولی راهکاری برای دستیابی به عدالت در ارزیابی انجام شد. ۱۶ روش تعیین حد نصاب قبولی یا نقطه‌ی برش مورد بررسی قرار گرفت. این روش‌ها شامل آنگوف، انگوف تغییر یافته، آنگوف/بله/خیر، آنگوف سه سطحی، آنگوف مبتنی بر نظریه‌ی سؤال پاسخ، بوکماریک، هافستی، نمره‌ی ثابت، رگرسیون مرزی، کوهن، ایبل، ندلسکی، گروه متضاد، روش جدید مبتنی بر معدل کل دانشجویان در سال قبل، نقشه سؤال، EZ می‌باشد. نهایتاً راهکارهای زیر به عنوان راهنمایی برای تصمیم‌گیران و مدیران آزمون‌ها و نیز اعضای هیأت علمی دانشگاه‌ها ارائه می‌گردد:

- استفاده‌ی ترکیبی از دو یا سه روش در تعیین حد نصاب قبولی برای دستیابی به دقت و اعتبار بیشتر.

- با توجه به مزایا و محدودیت‌های روش‌های مختلف پیشنهاد می‌شود در صورت ضرورت انتخاب یک روش برای تعیین حد نصاب قبولی، روش نقشه سؤال مبنا قرار گیرد. این روش با قرار دادن

مرتبط با سیستم آموزشی و دانشجو دارد برای ارزیابی دانشجویان از آزمون‌های مختلفی استفاده می‌شود. اگر آزمون را وسیله‌ای برای قضاوت در مورد هر یک از فراگیران تلقی کنیم، منطقی است نمره‌ای که مرز بین قبولی و مردودی فراگیران را تعیین می‌کند باید عادلانه و متناسب با شرایط آزمون باشد. به نمره‌ای که فراگیران را به دو دسته‌ی مردود و قبول تقسیم می‌کند؛ نمره‌ی برش (حد نصاب قبولی) آزمون می‌گویند. در واقع برای اطمینان از دستیابی دانشجویان به حداقل شایستگی، نیاز به تعیین معیاری برای قضاوت هر یک از آن‌ها داریم. از این جهت تعیین نمره‌ی برش امری ضروری است. برای محاسبه‌ی نقطه‌ی برش آزمون‌ها رویکردهای مختلفی وجود دارد هر چند که در اکثر مواقع از روش نمره‌ی ثابت استفاده می‌شود؛ مثلاً نقطه‌ی برش در آزمون‌های پیشرفت تحصیلی در مقطع کارشناسی ۱۲ و کارشناسی ارشد و دکتری برای دروس اختصاصی ۱۴ می‌باشد (نمره‌ی ثابت) و پایین‌تر از این نمره به منزله‌ی مردودی است.

از آنجا که هر ساله حدود ۵۰۰۰ پزشک فارغ‌التحصیل می‌شوند که انتظار می‌رود با دانش و مهارت فرار گرفته در دوران تحصیل واجد توانایی‌های لازم برای درمان و اداره‌ی بیماری‌های شایع جامعه باشند (۳۷). از طرفی در بسیاری از آزمون‌های مهم علوم پزشکی در ایران که منجر به کسب مدرک، اعطای مجوز و یا ارتقا به مراحل بالاتر می‌شود؛ دیده شده کسب ۵۰ یا ۷۰ درصد نمره‌ی کل آزمون به عنوان حد نصاب قبولی در نظر گرفته می‌شود. این ملاک گاهی باعث می‌شود برخی از دانشجویان در حالی به مقاطع بالاتر وارد شوند یا مدرک دانشگاهی کسب کنند که به بیش از ۵۰ درصد یا ۳۰ درصد اهداف آموزشی نرسیده‌اند. تصور یک پزشک یا متخصص با این میزان دانش و مهارت برای جامعه غیر قابل قبول و تنش‌زا است؛ همچنین اکثراً ملاکی که در این آزمون‌ها مورد استفاده قرار می‌گیرد بدون توجه به نوع و ماهیت سؤالاتی است که فرد درست پاسخ داده است در حالی که دانستن بخشی از مطالب در مقایسه با سایر مطالب از اهمیت بیشتری برخوردار است و عدم اطلاع از آن می‌تواند عواقب وخیم‌تری برای جامعه به دنبال داشته باشد. لذا در جهت مواجهه با مشکلات فوق می‌توان ساز و کارهای زیر را در فرایند تصمیم‌گیری در خصوص آزمون‌ها اتخاذ نمود.

در تعیین نقطه‌ی برش، گاه سؤالات آزمون و گاهی خود دانشجویان مورد بررسی قرار می‌گیرند. بعنوان مثال در روش آنگوف به طور کلی داوران به جای ارزیابی دانشجویان در مورد سؤالات آزمون نظر می‌دهند. روش آنگوف از این نظر که سؤالات آزمون را مورد ارزیابی قرار می‌دهد همانند روش بوکماریک است با این تفاوت که در بوکماریک لیست سؤالات آزمون همراه با ترتیب ضریب دشواری آن‌ها در اختیار داوران قرار داده می‌شود. برخلاف روش

برگزاری آزمون‌های متعدد در دانشگاه‌ها و وزارت متبوع، تنوع در شیوه‌های برگزاری آزمون و ترجیح اساتید، نیاز است که این موضوع به عنوان یکی از کارگاه‌های الزامی در واحد توانمندسازی اساتید ارایه گردد و علاوه بر آن تیمی ماهر و مسلط بر مفاهیم آزمون در سطح دانشگاه‌ها و نیز دبیرخانه‌های زیر مجموعه معاونت آموزشی وزارت بهداشت تشکیل شده تا مسئولیت نظارت و ارائه راهکارهای مؤثر جهت تعیین حد نصاب نمره قبولی برای هر یک از آزمون‌ها را به اعضای هیأت علمی دانشگاه‌ها و اعضای هیأت‌های بوردرشته‌های مختلف به عهده داشته باشد.

### تشکر و قدردانی

در پایان از تمامی افرادی که ما را در این مطالعه یاری نمودند؛ بالخصوص اساتید دانشگاه علوم پزشکی اصفهان، تشکر و قدردانی می‌نماییم.

توانایی داوطلب و دشواری سؤال در یک مقیاس مشترک، امکان محاسبه احتمال پاسخ صحیح را فراهم می‌کند. مزیت اصلی این روش، ارائه‌ی یک نمایش گرافیکی از رابطه بین توانایی داوطلب و دشواری سؤال است که به داوران کمک می‌کند تا حد نصاب قبولی را دقیق‌تر تعیین کنند. روش نقشه‌ی سؤال، با کاهش پیچیدگی شناختی و ارائه یک تصویر بصری از دشواری سؤالات، فرایند داوری را ساده‌تر می‌سازد. ویژگی‌های کلیدی این روش شامل مقیاس‌گذاری مشترک برای توانایی و دشواری سؤالات، تعیین حد نصاب بر اساس تحلیل داده‌ها، تفسیر آسان نتایج و ارائه‌ی اطلاعات عمیق در مورد عملکرد آزمون‌دهندگان است. به طور خلاصه، روش نقشه سؤال یک ابزار قدرتمند برای ارزیابی دقیق و عادلانه آزمون‌ها است.

-از آن جایی که بکارگیری مدیران کارآمد و لایق در راستای مدیریت تدوین نقشه جامع در راستای کیفیت آزمون‌ها می‌تواند در ارتقاء کیفیت آزمون مؤثر واقع شود (۳۹)؛ و از سوی دیگر با توجه به

### References

- 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 [in Persian]. Iranian Journal of Medical Education 2012; 11(8): 885-94.
- Habibzadeh S, Delavar A, Farrokhi N, Minaei A, Jalili M. The use of rasch and item mapping in determining cut score of comprehensive pre internship exam [in Persian]. Res Med Edu 2019; 11(3): 59-70.
- Makarem A, Mahdaviard H, Gholami H. Evaluation of Passing Scores in Semiotics: An Objective Structured Clinical Examination for Medical Students of Mashhad University of Medical Sciences, Mashhad, Iran, 2015. Strides Dev Med Educ 2017; 14(1): e59227.
- Deputy Ministry of Health. The regulations of the supreme council of planning in medical sciences. Educational regulations of the non-continuous master's course [in Persian]. 2010. Available from: <https://edu.mui.ac.ir/sites/edu/files/%D8%A2%DB%8C%DB%8C%D9%86%20%D9%86%D8%A7%D9%85%D9%87%20%DA%A9%D8%A7%D8%B1%D8%B4%D9%86%D8%A7%D8%B3%DB%8C%20%D8%A7%D8%B1%D8%B4%D8%AF.pdf>.
- Maghsoodi F, YazdanFard M, Pouladi S, Mirzaei K. Using the methods of fixed score, angoff, yes/no angoff and three-level angoff to determine the standard and acceptance rate of nursing students participating in an objective structured clinical exam [in Persian]. Journal of Medical Education 2023; 23: 217-24.
- Grabovsky I, Pace J, Runyon CJAPM. Cut-score operating function extensions: penalty-based errors and uncertainty in standard settings. Appl Psychol Meas 2021; 45(7-8): 536-50.
- Barman A. Standard setting in student assessment: is a defensible method yet to come? Ann Acad Med Singap 2008; 37(11): 957-63.
- Sam AH, Millar KR, Westacott R, Melville CR, Brown CA. Standard setting Very Short Answer Questions (VSAQs) relative to Single Best Answer Questions (SBAQs): does having access to the answers make a difference? BMC Med Educ 2022; 22(1): 640.
- Torabinia N, Razavi SM, Shayan S, Hatamzade Z, Reisi M. Assessment of the Isfahan, Yazd, Kashan students' views about the dentistry basic science exam [in Persian]. J Iranian Journal of Medical Education 2019; 19(0): 216-24.
- Nemat Bakhsh M. Are the grades of basic science courses related to the grades of the same courses in the comprehensive exam of basic sciences? [in Persian]. Iranian Journal of Medical Education 2004; 4(1): 82-5.
- Daryazadeh S, Shayan S. Overview of the evolution process of student assessment methods in medical education: report of the change process in Iran based on the 8-step model [in Persian]. Iranian Journal of Medical Education 2017; 17: 32-42.
- Deputy Ministry of Health. Executive instructions for registration of the 70th course of the specialized Board exam. Available from: <https://sanjeshp.ir/PDF.aspx?newsid=102501&type=application/pdf>.
- Kareshki H, Hajiabadi F, Bagheri M, Moghaddam AG. A review of the basics of determining the cut-off point in academic achievement tests and the introduction of Angoff scientific method [in Persian]. Horizon of Medical Education Development 2021; 12(4): 85-96.
- Taylor DD, Reid C, Senhauser DA, Shively JA. Use of minimum pass levels on pathology examinations. J Med Educ 1971; 46(10): 876-81.



15. Liu M, Liu KM. Descriptive terminology of education. *Kaohsiung J Med Sci* 2008; 24(12): 656-63.
16. Yousefi HR. Descriptive terminology of education [in Persian]. Tehran, Iran: Kalrj Bartar Publications; 2011.
17. Salzman DH, Rising KL, Cameron KA, Powell RE, Papanagnou D, Doty A, et al. Setting a minimum passing standard for the uncertainty communication checklist through patient and physician engagement. *J Grad Med Educ* 2020; 12(1): 58-65.
18. Ward H, Chiavaroli N, Fraser J, Mansfield K, Starmer D, Surmon L, et al. Standard setting in Australian medical schools. *BMC Med Educ* 2018; 18(1): 80.
19. Royal KD, Guskey TR. On the appropriateness of norm- and criterion-referenced assessments in medical education. *Ear Nose Throat J* 2015; 94(7): 252-4.
20. Alshawwa L. Standard setting: a review of methods. *Asian Journal of Education and Social Studies* 2023; 42(2): 1-7.
21. Kaufman DM, Mann KV, Muijtjens AM, van der Vleuten CPJAM. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Acad Med* 2000; 75(3): 267-71.
22. Park J, Ahn DS, Yim MK, Lee J. Comparison of standard-setting methods for the Korea Radiological technologist Licensing Examination: Angoff, Ebel, Bookmark, and Hofstee. *J Educ Eval Health Prof* 2018; 15: 32.
23. Shayan SH. Using Patient Management Problem (EPMP) in Assessment of Clinical Competency. *Iranian Journal of Medical Education* 2011; 10(5): 1087-92.
24. Dwyer T, Wright S, Kulasegaram KM, Theodoropoulos J, Chahal J, Wasserstein D, et al. How to set the bar in competency-based medical education: standard setting after an Objective Structured Clinical Examination (OSCE). *BMC Med Educ* 2016; 16: 1-7.
25. Jalalizadeh M, Delavar A, Farokhi N, Askari M. Comparison of ANGOF-based IRT method and Bookmark method for standard Setting of MSRT language test. *J Research in Teaching* 2019; 7(4): 69-49.
26. Yazdani S, Araghian F, Sanagoo A, Jooybari L. Hofstee an evaluation method in national medical science olympiad in Iran [in Persian]. *Iranian Journal of Medical Education* 2016; 16: 249-50.
27. Jalili M, Khabaz Mafinejad M, Gandhamkar R, Mortazhejri S. Principles and evaluation methods of learners in medical sciences [in Persian]. Tehran: Academy of Medical Sciences of the Islamic Republic of Iran; 2017: 629.
28. Taylor CA. Development of a modified Cohen method of standard setting. *Med Teach* 2011; 33(12): e678-e82.
29. Zieky M, Perie M. A primer on setting cut scores on tests of educational achievement. [2006] Available from: [https://www.ets.org/research/policy\\_research\\_reports/publications/publication/2006/dbkw.html](https://www.ets.org/research/policy_research_reports/publications/publication/2006/dbkw.html).
30. Bourque J, Skinner H, Dupré J, Bacchus M, Ainslie M, Ma IW, et al. Performance of the Ebel standard-setting method for the spring 2019 Royal College of Physicians and Surgeons of Canada internal medicine certification examination consisting of multiple-choice questions. *J Educ Eval Health Prof* 2020; 17: 12.
31. Livingston SA, Kastrinos WJ. A study of the reliability of Nedelsky's method for choosing a passing score. *ETS Research Report* 1982; 1982(1): i-13.
32. Shulruf B, Yang Y-Y, Huang P-H, Yang L-Y, Huang C-C, Huang C-C, et al. Standard setting made easy: validating the Equal Z-score (EZ) method for setting cut-score for clinical examinations. *BMC Med Educ* 2020; 20(1): 167.
33. Aung MN, Jaroovanichkul V, Deerojanawong J, Somboonwong J, Ishtiaq A, Wannakrairot PJ, et al. A New Method for Setting Standard in Medical Education, Applying Previous Year Cumulative GPA. 2019; 1(5).
34. Anbari K, Ahmadi SAY, Rezaian S, Sabooteh T, Shirzadehgan R, Pirzadroozbahani TJ. Using Cochran formula for investigation of the significance of students' passing or failing: an evidence-based protocol in medical education. *Iranian Journal of Medical Education* 2018; 18: 319-21.
35. Lahner F-M, Schaubert S, Lörwald AC, Kropf R, Guttormsen S, Fischer MR, et al. Measurement precision at the cut score in medical multiple choice exams: Theory matters. *Perspect Med Educ* 2020; 9(4): 220-8.
36. Kamali F, Shakour M, Yousefy A. Peer assessment in evaluation of medical sciences students [in Persian]. *Iranian Journal of Medical Education* 2012; 11(9): 1443-52.
37. Sabouri M, Shayan S, Salehi A sabori mosih ss, salehi ahmad. Reviewing and reorganizing the educational process of the medical internship courses of Isfahan University of Medical Sciences in 2013 (the preliminary stage of determining the educational needs of the women's department). *Iranian Journal of Education in Medical Sciences* 2002; 8: 48-9.
38. Kaftandjieva F. Methods for setting cut scores in criterion-referenced achievement tests. 2010; 170.
39. Kazemi M, Shayan S, Akbari Jor H. Determining and improvement the effective factors of board examination in internal medicine-General surgery-Pediatrics & gynecology (GYN) discipline from perspective of stakeholders (faculty member, residents and educational manager) [in Persian]. *Medicine and Spiritual Cultivation* 2019; 28(4): 41-51.

## Determining the Passing Score as A Strategy to Achieve Justice in Assessment: A Review Study

Atefeh Heidari<sup>1</sup>, Shahram Shayan<sup>2</sup>

### Review Article

#### Abstract

**Background:** One of the challenging issues in education is establishing fairness in student evaluation. Determining a fixed passing score for all tests without considering the specific characteristics of each test and test taker is unfair. The passing standard for each test should be determined based on factors such as the level of difficulty of the test, the ability of the students, the learning environment, and educational goals to establish fairness. This study was conducted to provide solutions for accurately determining the passing score threshold and improving the quality of tests. It is expected that the results of this study can help test administrators and university faculty members to improve the evaluation process and take an influential step toward establishing educational fairness.

**Methods:** This study was a narrative review and systematic search. Published Persian and English articles were searched in databases such as PubMed, Google Scholar, Scopus, and SID using keywords including standard setting, passing score, minimum pass level, and cut point. Finally, 37 studies were included in the study based on the inclusion and exclusion criteria.

**Findings:** The articles included in the study described 16 methods for determining the passing score. These were broadly divided into four categories: test-centered, examinee-centered, performance-centered, and outcome-centered.

**Conclusion:** To improve the quality of their exams, universities should monitor this process by utilizing experts in the evaluation and using mixed methods or item mapping as a standard approach to achieving fairness in assessments.

**Keywords:** Educational measurement/methods; Social justice; Medical education

**Citation:** Heidari A, Shayan Sh. **Determining the Passing Score as A Strategy to Achieve Justice in Assessment: A Review Study.** J Isfahan Med Sch 2025; 43(802): 53-62.

1- MSc Student, Department of Medical Education, Center for Studies and Development of Medical Education, Isfahan University of Medical Sciences, Isfahan, Iran.

2- Assistant Professor, Department of Medical Education, Center for Studies and Development of Medical Education, Isfahan University of Medical Sciences, Isfahan, Iran.

**Corresponding Author:** Shahram Shayan, Assistant Professor, Department of Medical Education, Center for Studies and Development of Medical Education, Isfahan University of Medical Sciences, Isfahan, Iran; Email: shahram\_shayan10@yahoo.com