

## مقایسه نمره حد نصاب و میزان قبولی شرکت کنندگان در آزمون عینی ساختارمند بالینی با استفاده از چهار روش مختلف تعیین استاندارد: نمره ثابت، آنگوف، رگرسیون مرزی و کوهن

محمد جلیلی<sup>۱</sup>، سارا مرتاض هجری<sup>۲\*</sup>

۱. متخصص طب اورژانس، دانشیار گروه طب اورژانس، مرکز تحقیقات آموزش علوم پزشکی، دانشگاه علوم پزشکی تهران، ایران  
۲. دانشجوی دکتری تخصصی آموزش پزشکی، مرکز مطالعات و توسعه آموزش، دانشگاه علوم پزشکی تهران، ایران

دریافت مقاله: ۹۰/۶/۱۲ • آخرین اصلاح مقاله: ۹۰/۹/۲ • پذیرش مقاله: ۹۱/۱/۲۰

**زمینه و هدف:** روش‌های مختلفی برای تعیین نمره حد نصاب قبولی آزمون‌ها وجود دارند اما تاکنون به ندرت برای امتحانات داخل کشور مورد استفاده قرار گرفته‌اند. ما در این مطالعه، حد نصاب قبولی آزمون عینی ساختارمند بالینی مقطع پیش کارورزی را به چهار روش نمره ثابت، آنگوف، رگرسیون مرزی و کوهن محاسبه کردیم.

**روش کار:** برای تعیین استاندارد به روش نمره ثابت، نمره ۶۰ (از ۱۰۰) در نظر گرفته شد. برای روش آنگوف، بعد از طراحی سؤالات، پانلی شامل ۱۱ داور تشکیل شد که هر یک به صورت مستقل، احتمال قبولی یک دانشجوی مرزی را در هر یک از ایستگاه‌ها برآورد کردند. میانگین احتمالات تمام داوران در تمام ایستگاه‌ها، استاندارد آزمون محسوب شد. به منظور اجرای روش رگرسیون مرزی برای هر ایستگاه، معادله رگرسیونی با نمره چک لیست و نمره ارزیابی کلی دانشجویان نوشته شد و استاندارد ایستگاه به ازای نمره ۲ در لیکرت محاسبه شد. در روش کوهن، ۶۰ درصد میانگین نمرات ۵ درصد اول دانشجویان در کل آزمون، به عنوان استاندارد آزمون در نظر گرفته شد.

**یافته‌ها:** استاندارد حاصل از روش‌های عدد ثابت، آنگوف، رگرسیون مرزی و کوهن، به ترتیب ۶۰، ۴۹/۱۵، ۴۲/۳۹ و ۴۲/۷۴ به دست آمد. میزان قبولی دانشجویان در کل آزمون به ترتیب ۲۱/۹، ۶۷/۶، ۹۳/۳ و ۹۲/۴ درصد بود.

**نتیجه گیری:** در این مطالعه، چهار روش برای تعیین استاندارد قبولی آزمون عینی ساختارمند بالینی مورد مطالعه و بررسی قرار گرفت و نشان داد روش اتخاذ شده برای تعیین حد نصاب قبولی آزمون، تاثیر قابل توجهی بر نمره استاندارد تعیین شده و درصد قبولی دانشجویان دارد.

**کلید واژه‌ها:** تعیین استاندارد، آزمون عینی ساختارمند بالینی، روش رگرسیون مرزی، روش آنگوف، روش کوهن

\* نویسنده مسؤول: دانشکده پزشکی، ساختمان شماره یک، اتاق ۱۱۲، بلوار کشاورز، خیابان قدس، خیابان پورسینا، تهران، ایران

## مقدمه

روش‌های مختلفی برای تعیین حد نصاب قبولی (استاندارد) آزمون وجود دارد (۳-۱). روش معمولی که برای اکثر امتحانات کشور به کار می‌رود، در نظر گرفتن نمره ثابت است که جدا از نوع آزمون و سختی آن برای تمام آزمون‌ها یکسان در نظر گرفته می‌شود. این نمره ثابت، در امتحانات پزشکی در مقطع علوم پایه عموماً عدد ۱۰ و در مقاطع بالینی، عدد ۱۲ است. علی‌رغم رایج بودن این روش، راه‌های گوناگون دیگری هم برای تعیین استاندارد آزمون وجود دارد اما در حد اطلاع نویسندگان تاکنون برای امتحانات سراسری و رسمی داخل کشور استفاده نشده است. یکی از روش‌هایی که نسبتاً زیاد استفاده می‌شود، روش آنگوف است که در آن، پانلی از متخصصان و داوران، معمولاً قبل از آزمون تشکیل می‌شود و از ایشان درخواست می‌شود تا بر اساس مدل‌لوزی خاصی، محتوای آزمون یعنی آیتم‌های مختلف آن را تحلیل و بررسی و استاندارد را تعیین کنند (۴). روش رگرسیون مرزی یکی دیگر از روش‌های تعیین استاندارد است که برای آزمون‌های عملکردی استفاده می‌شود و در آن نیازی به پانل متخصصان نیست بلکه در جریان یک آزمون واقعی، استاندارد مشخص می‌شود (۵). روش کوهن، که مشابه آن برای محاسبه حد نصاب آزمون‌های پزشکی جامع کشوری کاربرد داشته، احتساب درصد مشخصی (مثلاً ۶۰ درصد) از نمرات بالای آزمون است (۶).

جدا از اینکه روش تعیین استاندارد چه باشد، مدل نمره‌دهی یعنی اینکه چگونه نمرات دانشجویان در ایستگاه‌های مختلف را با هم ترکیب کنیم تا به نمره کلی برسیم، می‌تواند به صورت جبرانی یا غیرجبرانی در نظر گرفته شود. در حالت غیرجبرانی مهم است که دانشجویان در تک‌تک ایستگاه‌ها به حد نصاب قبولی برسند تا در مجموع آزمون قبول شود.

از سال ۱۳۸۷ تاکنون، از دانشجویان پزشکی دانشگاه علوم پزشکی تهران در بدو ورود به دوره پیش‌کارورزی، یک آزمون بالینی ساختارمند عینی (OSCE) به عمل می‌آید و قرار است

از سال ۱۳۹۰ یک آزمون مشابه در پایان دوره کارورزی انجام شود و صلاحیت بالینی دانشجویان قبل از فراغت از تحصیل مورد ارزیابی قرار گیرد. بدین منظور لازم است آزمون مورد نظر را یک آزمون معیار-محور با نمره حد نصاب قبولی در نظر بگیریم و بنابراین تعیین استاندارد آن حائز اهمیت است.

مطالعات مختلف نشان داده‌اند که روش‌های مختلف، استانداردها و درصد قبولی‌های متفاوتی به دست می‌دهند و نمی‌توان یکی از آنها را به عنوان بهترین روش توصیه کرد (۷،۸). این مقاله حاصل یک طرح تحقیقاتی است که توسط دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران تصویب شد و هدف از آن، تعیین استاندارد برای آزمون OSCE پیش‌کارورزی بود. در واقع ما نمره قبولی را با چند روش استخراج کردیم تا نقاط قوت و ضعف آنها را نسبت به یکدیگر مقایسه کنیم و ببینیم استفاده از هر کدام آنها در چه شرایطی باید انجام شود تا نتایج بهینه‌ای به دنبال داشته باشد. در این مقاله، قسمتی از نتایج طرح که مربوط به مقایسه چهار روش تعیین استاندارد است، ارائه می‌شود.

## روش کار

OSCE پیش‌کارورزی در تاریخ ۱۸ اسفند ۱۳۸۸ برگزار شد و ۱۰۵ دانشجوی پزشکی به صورت داوطلبانه در آن شرکت کردند. این آزمون شامل ۱۴ ایستگاه بود که دانشجویان در هر ایستگاه ۴ دقیقه فرصت داشتند تا وظیفه محوله را انجام دهند. از آزمونگر مستقر در هر ایستگاه خواسته شد تا علاوه بر پر کردن چک‌لیست و دادن ریز نمرات به یک مهارت، ارزیابی کلی خود را نسبت به عملکرد دانشجویان به صورت یک لیکرت ۵ تایی (رد، مرزی، قابل قبول، خوب، فراتر از انتظار) برآورد کند. برای یکسان کردن نمرات چک‌لیست در ایستگاه‌های مختلف، تمام نمرات از ۱۰۰ حساب شدند. تعیین استاندارد این آزمون، به چهار روش عدد ثابت، آنگوف، رگرسیون مرزی و کوهن انجام شد:

در لیکرت (دانشجوی مرزی)، چه نمره‌ای از چک لیست مورد انتظار است. به این ترتیب، استاندارد هر ایستگاه به دست آمد و برای محاسبه استاندارد کل آزمون، از آنها میانگین گرفته شد. روش کوهن: در این روش، ۶۰ درصد میانگین نمرات ۵ درصد اول دانشجویان در کل آزمون، به عنوان استاندارد آزمون در نظر گرفته شد.

کلیه داده‌ها اعم از سؤالات آزمون و اطلاعات شخصی دانشجویان و داوران محرمانه بود و فقط در اختیار افراد مسئول طرح قرار داده شد. استاندارد تعیین شده صرفاً کاربرد تحقیقاتی داشت و معیار قبولی یا ردی شرکت‌کنندگان در آزمون قرار نگرفت.

در این مطالعه، مدل نمره دهی جبرانی مورد استفاده قرار گرفت. پس از تعیین استاندارد کل آزمون بر اساس هر یک از روش‌های فوق، بر اساس توزیع نمرات دانشجویان در امتحان و میزان قبولی آنها تعیین شد. آنالیزها توسط نرم افزار SPSS.V.15 انجام شد.

### یافته‌ها

توزیع نمرات دانشجویان در ایستگاه‌های مختلف در جدول ۱ آمده است.

نمره ثابت: از آنجا که OSCE پیش‌کارورزی، یک آزمون در مقطع بالینی است، حدنصاب آن را در این مطالعه مشابه نمره قبولی در هر یک از دروس و بخش‌های بالینی ۱۲ (۶۰٪ از ۱۰۰) در نظر گرفتیم.

روش آنگوف: در این روش، پس از طراحی سؤالات توسط کمیته آزمون، تعدادی از اعضای هیأت علمی دانشکده پزشکی که با کارآموزان سال‌های چهارم و پنجم پزشکی کار کرده بودند و در زمینه برگزاری OSCE تجربه داشتند، با در نظر گرفتن بیشترین تنوع از لحاظ تخصص، سابقه آموزشی، سن و جنس برای شرکت در کارگاه آنگوف دعوت شدند. در این کارگاه که با حضور ۱۱ نفر از اعضای هیأت علمی برگزار شد، بعد از معرفی روش، از هر داور خواسته شد تا به صورت انفرادی هر ایستگاه را بررسی کند و احتمال قبول شدن یک دانشجوی مرزی در ایستگاه مورد نظر را به صورت مقداری از ۰ تا ۱۰۰ بیان کند. به این ترتیب هر داور یک عدد برای هر ایستگاه بیان کرد. میانگین نظر داوران برای یک ایستگاه، استاندارد آن ایستگاه در نظر گرفته شد و میانگین استاندارد ایستگاه‌های مختلف، استاندارد کل آزمون محسوب شد (۹).

روش رگرسیون مرزی: در این روش نمره چک لیست دانشجویان در هر ایستگاه به عنوان متغیر وابسته و نمره حاصل از ارزیابی کلی (در مقیاس لیکرت) به عنوان متغیر مستقل وارد یک معادله رگرسیون شدند و محاسبه شد که به ازای نمره ۲

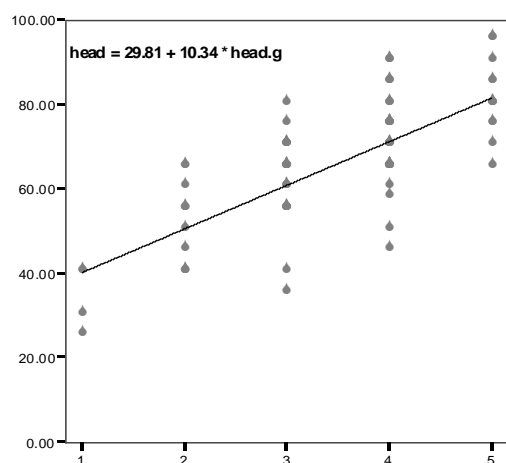
جدول ۱: توزیع نمرات دانشجویان شرکت‌کننده در OSCE پیش‌کارورزی به تفکیک ایستگاه

شماره ایستگاه	نام ایستگاه	پایین ترین نمره	بالا ترین نمره	میانگین	انحراف معیار
۱	شرح حال سردرد	۲۵	۹۵	۶۶/۳۶	۱۳/۸۰
۲	مهارت ارتباطی	۲۶	۹۶	۶۲/۵۹	۱۶/۶۰
۳	معاینه قلب و ریه	۰	۷۱/۴۳	۲۸/۱۶	۱۹/۶۸
۴	معاینه پستان	۳۵/۲۷	۱۰۰	۷۹/۶۷	۱۳/۹۲
۵	اداره سوختگی	۵	۱۰۰	۳۴/۷۶	۱۸/۳۹
۶	تشخیص ضایعه پوستی	۰	۹۰	۴۸/۱۴	۲۶/۸۷
۷	معاینه ژنیتال	۰	۱۰۰	۵۱/۳۱	۲۴/۴۵
۸	آتل‌گیری	۰	۱۰۰	۱۱/۲۴	۱۹/۲۵

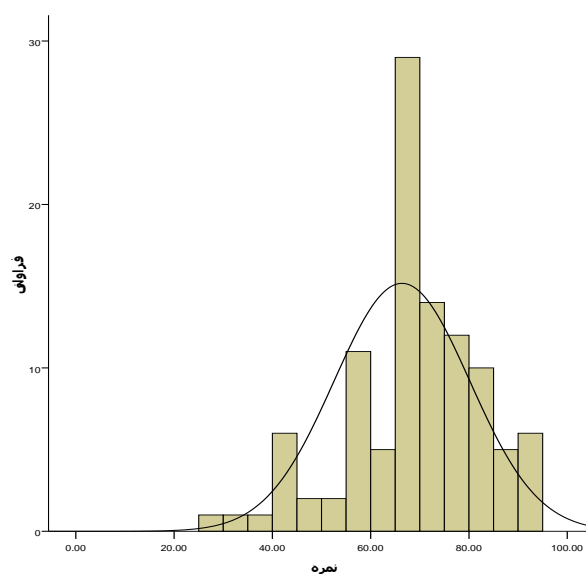
۱۶/۹۵	۷۷/۸۲	۱۰۰	۲۸/۵۷	انجام سونداژ	۹
۱۵/۵۸	۶۵/۱۵	۹۵/۸۳	۸/۳۳	معاینه شکم	۱۰
۲۱/۰۸	۷۸/۵۰	۱۰۰	۰	بخیه زدن	۱۱
۳۳/۳۸	۴۸/۱۰	۱۰۰	۰	آزمون تیروئید	۱۲
۲۱/۲۹	۶۰/۴۶	۱۰۰	۰	رفلکس نوزادی	۱۳
۱۷/۳۳	۵۴/۲۵	۹۶	۲۰	شرح حال اسهال	۱۴
۸/۸۰	۵۴/۱۱	۷۵/۶۵	۳۱/۴۵	کل آزمون	

نوشته شد که با قرار دادن نمره ۲ (عملکرد دانشجوی مرزی در لیکرت) استاندارد هر ایستگاه محاسبه شد. به عنوان نمونه، نمودار رگرسیون مربوط به ایستگاه شرح حال در نمودار ۲ نشان داده شده است (۹).

بیشترین و کمترین میانگین نمرات دانشجویان به ترتیب مربوط به ایستگاه‌های معاینه پستان (۷۹/۶۷) و ایستگاه آتل‌گیری (۱۱/۲۴) بود. به عنوان نمونه، توزیع نمرات دانشجویان در ایستگاه شرح حال در نمودار ۱ نشان داده شده است (۹).



نمودار ۲: توزیع نمرات چک لیست دانشجویان در ایستگاه شرح حال بیمار مبتلا به سردرد نسبت به نمره گلوبال ایشان در همین ایستگاه



نمودار ۱: توزیع نمرات دانشجویان در ایستگاه شرح حال بیمار مبتلا به سردرد در OSCE پیش کارورزی ۱۳۸۸

میزان قبولی دانشجویان بر اساس این روش نیز تعیین شد که در جدول ۲ نشان داده شده است، قابل مشاهده می‌باشد. بر اساس مدل نمره‌دهی جبرانی، استاندارد کل آزمون بر اساس روش‌های عدد ثابت، آنگوف، رگرسیون مرزی و کوهن به ترتیب ۶۰، ۴۹/۱۵، ۴۲/۳۹ و ۴۲/۷۴ محاسبه شد. بر اساس استانداردهای فوق، میزان قبولی دانشجویان در کل آزمون در

در روش آنگوف، ابتدا استاندارد تک تک ایستگاه‌ها محاسبه شد و بر اساس آن، میزان قبولی دانشجویان در هر ایستگاه تعیین شد. در روش رگرسیون مرزی، در هر ایستگاه بر اساس نمرات چک لیست و نمرات گلوبال، معادله رگرسیونی

روش‌های عدد ثابت، آنگوف، رگرسیون مرزی و کوهن به ترتیب ۲۱/۹، ۶۷/۶، ۹۳/۳ و ۹۲/۴ درصد به دست آمد.

جدول ۲: استانداردها و میزان قبولی تعیین شده بر اساس روش‌های آنگوف و رگرسیون مرزی، به تفکیک هر ایستگاه و کل آزمون OSCE پیش‌کاروری

شماره ایستگاه	نام ایستگاه	روش آنگوف		رگرسیون مرزی	
		میزان قبولی (درصد)	استاندارد (از ۱۰۰)	میزان قبولی (درصد)	استاندارد (از ۱۰۰)
۱	شرح حال سردرد	۵۸.۶۴	۷۷/۱	۸۷/۶	۵۰.۴۹
۲	مهارت ارتباطی	۵۷.۲۷	۶۱/۰	۸۴/۸	۴۵.۴۱
۳	معاینه قلب و ریه	۳۹.۵۵	۳۴/۳	۵۸/۱	۲۷.۰۷
۴	معاینه پستان	۴۳.۶۴	۹۹/۰	۸۹/۵	۶۱.۴۴
۵	اداره سوختگی	۳۹.۰۹	۴۹/۵	۵۶/۲	۳۰.۷۶
۶	تشخیص ضایعه پوستی	۴۵.۹۱	۶۳/۸	۶۶/۷	۳۸.۳۱
۷	معاینه ژنیتال	۴۵.۰۰	۶۱/۹	۶۱/۹	۴۶.۸۱
۸	آتل‌گیری	۴۰.۰۰	۱۱/۴	۱۹/۰	۲۹.۷۳
۹	سونداژ	۶۳.۱۸	۸۳/۸	۸۴/۸	۵۷.۴۰
۱۰	معاینه شکم	۵۰.۹۱	۶۱/۰	۷۷/۱	۴۲.۷۸
۱۱	بخیه زدن	۵۲.۷۳	۸۷/۶	۸۷/۶	۵۳.۹۴
۱۲	آزمون تیروئید	۴۹.۵۵	۶۵/۷	۷۶/۲	۲۴.۱۵
۱۳	رفلکس نوزادی	۴۵.۹۱	۷۰/۵	۷۴/۳	۴۳.۶۴
۱۴	شرح حال اسهال	۵۶.۸۲	۳۷/۱	۷۳/۳	۴۱.۵۲
	کل آزمون	۴۹/۱۵	۶۷/۶	۹۳/۳	۴۲/۳۹

## بحث و نتیجه‌گیری

ما در این مطالعه چهار روش عدد ثابت، آنگوف، رگرسیون مرزی و کوهن را مورد بررسی قرار دادیم. هرچند برای تعیین استاندارد آزمون‌های مبتنی بر عملکرد مطالعات زیادی انجام شده اما به عقیده پژوهش‌گران هیچ روش تعیین استاندارد برای OSCE کامل و بی‌نقص نیست (۷،۸)؛ از طرفی استفاده از روش‌های مختلف، موجب تعیین استانداردهای مختلف می‌شود (۱۰،۱۱) و روش استاندارد طلایی برای مقایسه روش‌های تعیین استاندارد وجود ندارد. در برخی از مطالعات از میزان سخت‌گیرانه بودن یا آسان‌گیرانه بودن استانداردها، جهت مقایسه آنها توصیه شده است (۱،۱۲). در واقع این موضوع مشخص می‌کند که درصد قبولی‌رُدی در آزمون تا چه اندازه از نظر کادر آموزشی و اجرایی دانشکده، با توجه به

عملکرد این دانشجویان در سایر آزمون‌ها و همچنین در مقایسه با سایر مطالعات کوهورت دیگر از دانشجویان، پذیرفتنی و منطقی به نظر می‌رسد. در مطالعه ما بیشترین کمترین درصد

قبولی به ترتیب بر اساس رگرسیون مرزی و عدد ثابت به دست آمد که تفاوت بین آنها قابل توجه بود.

در روش نمره ثابت، استاندارد آزمون از قبل مشخص است و در واقع اساساً لازم نیست کار خاصی صورت گیرد. اما مشکل اینجاست که سختی آزمون و سطح داوطلبان را در نظر نمی‌گیرد. همچنین از آنجا که یک روش معیار-محور است و با عملکرد دانشجویان تغییر نمی‌کند، میزان قبولی دانشجویان در

تعیین استاندارد با روش رگرسیون مرزی نیز برای اولین بار در دانشکده پزشکی انجام شد. بنابراین قبل از شروع آزمون، در خصوص نمره دهی لیکرت با برگزارکنندگان صحبت شد. مسأله مهم این است که در هنگام ارزیابی کلی عملکرد دانشجویان در ایستگاه، باید سطح و مقطع دانشجویانی که در امتحان شرکت می‌کنند، مد نظر باشد و دانشجویان در همان سطح خودشان سنجیده شوند. همچنین نمره لیکرت، واقعاً باید منعکس کننده عملکرد کلی دانشجوی باشد و آزمونگر نباید نمره چک لیستی که دانشجو در ایستگاه کسب کرده را به نمره لیکرت تبدیل کند (۱۶).

در مطالعه اخیر، استاندارد که از روش مرزی به دست آمد نسبت به همه استانداردهای دیگر آسان‌گیرانه‌تر بود. در مطالعه کافمن (Kaufman) که پنج روش را با هم مقایسه کرده بود، استاندارد حاصل از روش مرزی و روش آنگوف مشابه بودند نویسندگان ادعا کردند که هر دوی این روش‌ها قابل دفاع و منطقی‌اند (۱۰). در مطالعه کرامر (Kramer) و همکاران، روش‌های رگرسیون مرزی و آنگوف انفرادی و آنگوف با بررسی نمرات واقعی با هم مقایسه شدند که استاندارد روش مرزی از همه کمتر بود. نویسندگان ذکر کردند که اختلاف بین استانداردهای حاصل از روش‌های مختلف آن قدر زیاد بود که عملاً نمی‌شد گفت کدام بهتر است. اما در کل عنوان کردند که استفاده از روش مرزی، علاوه بر اینکه آسان‌گیرتر است، روایی و پایایی بهتری به دست می‌دهد و مخصوصاً برای آزمون‌های مبتنی بر عملکرد بر آنگوف ارجح است (۵). در مطالعه دیگری، سه روش مرزی، آنگوف بلی/خیر و آنگوف تغییر یافته با هم مقایسه شدند. از نظر میزان قبولی، روش مرزی بیشتر از روش آنگوف و حتی آنگوف تغییر یافته بود. بنابراین در این مطالعه هم، روش مرزی سخت‌گیری کمتری نشان داد. به نظر نویسندگان نتایج حاصل از روش مرزی قابل قبول‌تر بود (۱۲).

یکی از محدودیت‌های مطالعه این بود که تعیین استاندارد با روش آنگوف برای اولین بار در دانشکده پزشکی انجام می‌شد و هیچ یک از داوران تجربه قبلی در این زمینه نداشتند و اساساً

امتحانات مختلف ممکن است متغیر باشد. مطالعه کوهن مشخص کرد که روش نمره ثابت، باعث شده میزان قبولی طی سال‌های مختلف، نوسان زیادی داشته باشد (از ۳ تا ۸۳ درصد) (۶).

محاسبه استاندارد در روش کوهن بسیار ساده است اما از آنجا که نمره دانشجویان برتر را ملاک قرار می‌دهد، استاندارد به طیف شرکت‌کنندگان بستگی دارد. البته یک روش مبتنی بر هنجار نیز محسوب نمی‌شود چون در عین حال، ملاک ۶۰ درصد را نیز در نظر می‌گیرد. در هر حال بر اساس نتایج مطالعه ما، میزان قبولی در روش نمره ثابت، به مراتب کمتر از روش کوهن بود. این یافته با مطالعه کوهن مطابقت دارد که با بررسی امتحانات در دو دانشگاه هلند طی ۹ سال نشان داد میانگین میزان قبولی طی این سال‌ها از روش نمره ثابت (۶۰) و روش کوهن به ترتیب ۴۷ و ۸۳ درصد بوده است (۶).

تعیین استاندارد به روش آنگوف از آنجا که در جلسه ای قبل از آزمون صورت می‌گیرد و نیاز به حضور هیأت علمی دارد، ممکن است همه جا قابل اجرا نباشد. گذشته از این، کاری که در جلسه آنگوف صورت می‌گیرد از لحاظ ذهنی پیچیده و دشوار است که ممکن است مشکل ایجاد کند. مخصوصاً در مطالعه حاضر که هیچ یک از داوران تجربه قبلی در این زمینه نداشتند. گفته شده که فضای مجازی و فرضی حاکم بر جلسه می‌تواند درک داور را از عملکرد دانشجوی مرزی تحت تأثیر قرار دهد به طوری که استاندارد بیش از حد بالا و سخت‌گیرانه تعیین شود (۱۰، ۱۳). یکی از راه‌حل‌های عنوان شده، اهمیت دادن به انتخاب داورها و همچنین آموزش دادن به آنهاست (۸، ۱۲، ۱۴، ۱۵). در مطالعه ما در ابتدای کارگاه آنگوف، توضیحاتی در مورد روش برای داوران ارائه شد و سپس از آنان خواسته شد در رابطه با روش، نحوه نمره دادن و مفهوم دانشجوی مرزی با یکدیگر بحث کنند تا به دید مشترک برسند. همچنین بعد از اینکه به صورت تمرینی استاندارد دو ایستگاه از آزمون سال قبل را تعیین کردند و بازخورد گرفتند، فرایند تعیین استاندارد را شروع کردند.

کمک‌کننده است. همچنین باید توجه داشت که چنانچه استفاده از روش رگرسیون مرزی مورد نظر باشد، اهمیت استفاده از آزمونگران متخصص و آزموده در ایستگاه‌ها که سابقه کار با دانشجویان را داشته باشند، اهمیت بسیاری دارد.

به طور کلی، نتایج این مطالعه نشان می‌دهد که روش اتخاذ شده برای تعیین حد نصاب قبولی آزمون، تاثیر قابل توجهی بر نمره استاندارد تعیین شده و درصد قبولی دانشجویان دارد. در مورد آزمون‌های معیار-محور، انتخاب روشی که از بیشترین میزان روایی و پایایی برخوردار باشد، ضروری است.

### سپاسگزاری

از زحمات جناب آقای دکتر لباغ، رییس محترم مرکز مهارت‌های بالینی و جناب آقای دکتر میرزازاده، مدیر محترم دفتر توسعه آموزش دانشکده پزشکی که در برگزاری آزمون و اجرای روش رگرسیون مرزی و آنگوف نقش عمده‌ای داشتند، سپاسگزاری می‌شود.

همچنین لازم به ذکر است که این مقاله حاصل بخشی از پایان‌نامه تحت عنوان "تعیین استاندارد آزمون عینی ساختارمند بالینی به سه روش آنگوف تغییر یافته، آنگوف سه سطحی و رگرسیون مرزی" در مقطع کارشناسی ارشد آموزش پزشکی می‌باشد که با حمایت دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران اجرا شده است.

### References:

1. Norcini JJ. Setting standards on educational tests. *Medical Education* 2003;37:464-469
2. Cizek GJ, Bunch MB. Standard setting: A guide to establishing and evaluating performance standards for tests. Thousand Oaks, CA: Sage Publications, Inc. 2007: 20-22.
3. Min Liu, Keh-Min Liu. Setting pass scores for clinical skills assessment. *Kaohsiung Journal of Medical Sciences* 2008;24(12):656-663
4. Angoff WH. Scales, norms and equivalent scores. In: Thorndike RL, ed. *Educational Measurement*. 2nd ed. Washington, DC: American Council on Education 1971:508-600
5. Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. *Objective structured clinical examinations. Medical Education* 2003; 2: 132-139.

مفاهیمی مانند عملکرد مرزی برای آنان جدید بود. یکی از راه‌حل‌های عنوان شده در متون، اهمیت دادن به انتخاب داورها و همچنین آموزش دادن به آنهاست. همانطور که ذکر شد در مطالعه ما در ابتدای کارگاه، توضیحاتی در مورد روش برای داوران ارائه شد و سپس از آنان خواسته شد در رابطه با روش، نحوه نمره دادن و مفهوم دانشجوی مرزی با یکدیگر بحث کنند تا به دید مشترک برسند (۹).

در مورد روش مرزی نیز مطالعه با محدودیت‌هایی مواجه بود. یکی اینکه آزمون‌گران، در تمام موارد اعضای هیأت علمی نبودند و به دلیل شرایط و امکانات برگزاری OSCE، در برخی از ایستگاه‌ها پزشکان عمومی و دستیاران تخصصی، عملکرد دانشجویان را مشاهده و ارزیابی می‌کردند. ممکن است این موضوع موجب این باعث سهل‌گیری شده باشد. از طرفی باز هم به خاطر شرایط برگزاری آزمون، امکان برگزاری جلسه‌های آشنایی با روش رگرسیون مرزی، نحوه نمره‌دهی به صورت گلوبال و لیکرت و توضیح و بحث در مورد خصوصیات دانشجوی مرزی وجود نداشت و آزمون‌گران تنها به صورت خلاصه و اجمالی، دقایقی قبل از آزمون در مورد نمره‌دهی گلوبال توجیه شدند.

برای انجام مطالعات بعدی به نظر می‌رسد کسب نظر از دست‌اندرکاران اجرایی آموزش دانشجویان در خصوص میزان واقعی بودن استانداردهای به دست آمده در تفسیر نتایج روش‌های مختلف، مفید است. همچنین به نظر می‌رسد راه‌اندازی نهادی اجرایی که در کنار کمیته برگزاری آزمون، مسوولیت تعیین استاندارد آزمون را بر عهده داشته باشد،

6. Cohen-Schotanus J, Vleuten Cees van der. A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical Teacher* 2010; 32: 154–160
7. Chesser A, Laing M, Miedzybrodzka Z, et al. Factor analysis can be a useful standard setting tool in a high stakes OSCE assessment. *Medical Education* 2004; 38: 825–831
8. Davison I, Bullock A. Evaluation of the Introduction of the Objective Structured Public Health Examination. The Research Office School of Education University of Birmingham 2007
9. 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):885-894. [In Persian]
10. Kaufman DM, Mann KV, Muijtjens AMM, van der Vleuten CPM. A comparison of standard setting procedures for an OSCE in undergraduate medical education. *Academic Medicine* 2000;75:267–71.
11. Cusimono MD, Rothman AI. The Effect of Incorporating Normative Data into a Criterion-Referenced Standard Setting in Medical Education. *Academic Medicine* 2003;78(10 suppl):88-90
12. Schoonheim-Klein M, Muijtjens A, Habets L and et al. Who will pass the dental OSCE? Comparison of the Angoff and the borderline regression standard setting method. *European Journal of Dental Education* 2009;13:162–171
13. Norcini JJ, Stillman PL, Sutnick AI, et al. Scoring and Standard Setting with Standardized Patients. *Evaluation & the Health Professions* 1993;16(3):322-332.
14. Yudkowsky R, Downing SM, Popescu Mihaela. Setting standards for performance tests: A pilot study of a three-level Angoff method. *Academic Medicine* 2008; 83 (10 suppl):13-16
15. Kane MT, Crooks TJ, Cohen AS. Designing and evaluating standard-setting procedures for licensure and certification tests. *Advanced Health Science Education Theory Practice* 1999;4:195–207
16. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics. *AMEE guide no. 49. Medical teacher* 2010;32:802–811



## Standard Setting for Objective Structured Clinical Exam Using Four Methods: Pre-fixed score, Angoff, Borderline Regression and Cohen's

*Mohammad Jalili<sup>1</sup>, Sara Mortazhejri<sup>2\*</sup>*

*1. Specialist in Emergency Medicine, Associate Professor Department of Emergency Medicine, Center for Educational Research in Medical Sciences, Tehran University of Medical Sciences, Iran*

*2. Ph.D. Candidate of Medical education, Educational Development Center, Tehran University of Medical Sciences, Iran*

• Received: 28 Aug, 2011

• Received Corrected Version: 23 Nov, 2011

• Accepted: 8 Apr, 2012

**Background & Objectives:** Different methods have been introduced to determine the cut-scores of exams. However, they have been rarely used in our country. In this study, we used four methods to set a standard for pre-internship objective structured clinical examination (OSCE) in Tehran University of Medical Sciences.

**Methods:** A pre-fixed score of 60% was considered in the first approach. In the Angoff method, a panel including 11 faculty members was invited to estimate the probability of passing each station for a borderline student and mean of all estimations was considered as the station standard. In borderline regression method, a linear regression model for each station was calculated in which the students' checklist scores and general scores were considered as dependent and independent variables respectively. The station standard was calculated based on grade 2 in Likert scale. In Cohen's method, we set the standard at 60% of the 95th percentile point.

**Results:** The standards for the total test in the pre-fixed score, Angoff, borderline regression and Cohen's methods were respectively 60, 49.15, 42.39 and 42.74. According to these standards, the percentages of passing students were respectively 21.9%, 67.6%, 93.3% and 92.4%.

**Conclusion:** Using four methods to set standard of an OSCE showed that this approach for determining exam's cut-score has significant effect on the calculated standard and the percent of students passing the exam.

**Key Words:** Standard setting, Objective structured clinical examination, Angoff, Borderline regression method, Cohen method

\*Correspondence: School of Medicine, Department No. 1, Room 112, Keshavarz Blvd, Ghods St, Poorsina St, Tehran Iran

• Tel: 021-64053365

• Fax: 021-88973670

• Email: sa\_mortazhejri@razi.tums.ac.ir