# تحلیل کیفیت آزمون عینی ساختارمند بالینی در دانشگاه علوم پزشکی البرز

سلیمان احمدی[۱]، امین حبیبی[۲]، میترا رحیم زاده[۳]، شهلا بهرامی[۴*]

[۱] گروه آموزش پزشکی، دانشکده مجازی آموزش پزشکی ، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
[۲] گروه آموزش پزشکی، دانشگاه هوشمند علوم پزشکی ، تهران، ایران
[۳] مرکز تحقیقات عوامل اجتماعی مؤثر بر سلامت،گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی البرز، کرج، ایران
[۴] دانشجوی کارشناسی ارشدآموزش پزشکی، دانشکده مجازی آموزش پزشکی ، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

تاریخ دریافت مقاله: ۱۴۰۱/۱۰/۲۰ ؛ تاریخ پذیرش: ۱۴۰۲/۶/۲۷

## چکیده

**مقدمه:** یکی از روشهای ارزشیابی مبتنی بر عملکرد، آزمون های عینی ساختارمند می باشد. گذراندن این آزمون جهت فراغت از تحصیل دانشجویان پزشکی، انجام طبابت و شرکت در دوره تخصصی ضروری است. این مطالعه به منظور تحلیل کیفیت این آزمون ها با استفاده از شاخص های سایکومتریک طراحی شده است.

**روش‌ها :** در این مطالعه توصیفی مقطعی آزمون های سه دوره صلاحیت بالینی ۳۱ و ۳۲ و ۳۳ مورد تحلیل قرار گرفتند. شاخص های میانگین (انحراف معیار)، نمره حدنصاب (به روش رگرسیون مرزی)، میزان ردی، ضریب تعیین $R^2$ ، ریشه میانگین مربعات خطا برای تخمین میزان خطای حد نصاب همچنین ضریب دشواری و ضریب تمیز در تک تک ایستگاه ها و کل آزمون ها محاسبه شد.

**نتایج:** ۶۴ دانشجوی پزشکی سال آخر پزشکی به همراه ۲۶ نفر آزمونگر متخصص هر ایستگاه در این آزمون ها حضور داشتند. با توجه به نمره حد نصاب رگرسیون مرزی در آزمون ۳۱ ام همه دانشجویان نمره قبولی را کسب کرده و مردود نداشتیم. RMSE کل معادل ۰/۵۵ بود و مقدار $R^2$ در ایستگاه های مختلف از ۰/۴۵تا ۰/۸۵ بدست آمد. سخت ترین و آسانترین ایستگاهها براساس ضریب دشواری به ترتیب زنان، پزشکی اجتماعی و ارتوپدی بود. در آزمون ۳۲ام میزان ردی یک نفر(۴/۳درصد) بود. RMSE کل ۰/۵۷ و مقدار $R^2$ در ایستگاه های مختلف از ۰/۳۸تا ۰/۸۷ بدست آمد. سخت ترین و آسانترین ایستگاهها بر اساس ضریب دشواری به ترتیب ایستگاه رادیولوژی، طب اورژانس و جراحی بود. در آزمون ۳۳ام همه دانشجویان نمره قبولی را کسب کردند. RMSE کل معادل ۰/۷۷ بود و مقدار $R^2$ در ایستگاه های مختلف از ۰/۳۹ تا ۰/۹۵ بدست آمد. سخت ترین و آسانترین ایستگاهها بر اساس ضریب دشواری به ترتیب داخلی، غدد و جراحی بود.

**نتیجه‌گیری:** برای ارتقای کیفیت آزمون OSCE سنجش کیفیت آن حیاتی است. استفاده از شاخص‌های سایکومتریک می تواند گامی مهم در تعیین کیفیت آزمون باشد. نتایج به دست آمده در طول سه دوره آزمون حاکی از آن است که در بیشتر ایستگاه ها کیفیت آزمون قابل قبول بوده و در چند مورد نیز نیازمند بازنگری و بهبود وضعیت هستند.

**واژه‌های کلیدی:** تحلیل آزمون، آزمون ساختارمند بالینی، صلاحیت بالینی، سایکومتریک، رگرسیون مرزی

***نویسنده مسئول:**
دانشجوی کارشناسی ارشد آموزش پزشکی، دانشکده مجازی آموزش پزشکی ، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

۰۹۹۰۱۸۶۸۷۰۷
bahrami3636@gmail.com



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

## مقدمه

آموزش به عنوان ثروتی ملی در گنجینه پیشرفت و ارتقای جوامع بشری به حساب می آید. در این راستا ارائه آموزشی صحیح و منطبق با نیاز جامعه و فراگیران، می تواند تنها راه دستیابی به سطوح مطلوب یادگیری و ارتقای سطح رضایت مندی فراگیران باشد.

یادگیری به عنوان یک موضوع مهم و پیچیده در فرایند زندگی انسان، قادر به ایجاد تغییر در دانش، نگرش، مهارتها، ارزشها، هیجانات، عقاید و احساسات می باشد [۱]. بر این اساس می توان گفت: قلمرو یادگیری گسترده بوده و زمینه های شناختی، عاطفی و روان حرکتی را شامل می شود[۲].

محیط یادگیری بالینی، محیطی است که دانشجویان براساس نیازهای بیمار، مهارتهای خود را با هدف حفظ و ارتقای سلامت بیمار، حمایت فیزیکی، روانی، معنوی و اجتماعی با یک رویکرد کلی به کار می بندند [۴]. در واقع در فرایند آموزش حرفه های علوم پزشکی، محیط بالینی حائز اهمیت بسیاری است [۵و۶] و نقش مهمی در کسب تجارب مثبت یا منفی دارد [۷].

گذر از نقش دانشجویی به نقش حرفه ای، به عنوان یک مرحله مهم در طول دوره شغلی پزشکان، از اهمیت بسیار زیادی برخوردار است. با داشتن تجربیات مطلوب در محیط‌های بالینی و عرصه‌های کاری، رفته رفته تجربیات حرفه‌ای در دانشجویان شکل می گیرد و به آنها فرصت به کارگیری یادگرفته‌های تئوریکی در دنیای واقعی کار و هماهنگ سازی تصورات با واقعیات را خواهد داد [۸]. همان قدر که مناسب و مطلوب بودن این محیط‌ها میتواند تاثیرات مثبتی به همراه داشته باشد، یک محیط یادگیری بالینی نامطلوب میتواند اثرات سویی را ایجاد نماید، چرا که اگر در اولین برخورد با عرصه‌های واقعی کار و محیط‌های بالینی، اضطراب بر دانشجویان غلبه یابد و ماندگار باقی بماند، آنها منفعل گشته و قدرت درک و یادگیری خود را از دست می‌دهند [۹]. محققان دریافته‌اند که بسیاری از تجارب منفی دانشجویان از عرصه‌های بالینی به سازماندهی ضعیف و ارتباط نامطلوب دانشگاه با بالین مربوط می شود [۱۰].

این نامطلوب بودن و عدم جذابیت عرصه های کاری می‌تواند علاقه دانشجویان به حرفه شان را کاهش دهد [۱۱]. برخی یافته‌ها نشان میدهد، محیط های بالینی از دیدگاه دانشجویان، به گونه‌ای نبوده که آنها را برای درک مفهوم مراقبت به عنوان یک فرایند تربیت نماید. تحقیقات کشور ما نیز نشان میدهد که دانشجویان محیط‌های بالینی را بسیار پراسترس تجربه نموده‌اند [۱۲و۱۳و۱۴].

مطالعات نشان داده‌اند که پزشکان دانش‌آموخته برخی مهارت‌های عملی و ضروری را ندیده‌اند و یا بطور مستقل انجام نداده‌اند و این امر موجب مشکلات بسیاری در اداره بیماران گردیده است. از جمله، در مطالعه‌ای، ۶۵درصد از دانش‌آموختگان، دانش کسب شده در دوران تحصیل را بر اساس جامعه‌نگری در حد متوسط، ۲۰ درصد در حد خوب و ۱۵ درصد ضعیف دانسته اند [۲].

ارزیابی قلب هر موسسه آموزشی است. در واقع این ارزیابی هست که یادگیری را هدایت می کند[۱۶]. از طریق ارزیابی می توان میزان یادگیری به دست آمده توسط کارآموز را اندازه گیری کرد و میزان اثر بخشی برنامه ها را بررسی کرد [۱۷]. یکی ازموضوعات اساسی در آموزش ارزیابی دانشجویان پزشکی ارتقای دانش، نگرش و مهارتهای لازم جهت طبابت و نیز ارزیابی صحیح و دقیق آنهاست. بدیهی است آموزش و ارزیابی بایستی به گونه ای باشد که دانشجوی پزشکی بتواند بعد از فراغت از تحصیل در تطابق کامل با محیط واقعی حرفه ای توانایی تطبیق با شرایط کاری را داشته باشد. بنابراین آموزش و ارزیابی بصورت سنتی یعنی تدریس شفاهی و آزمون کتبی قادر به تشخیص واقعی توانایی های فردی و کیفیت و کارایی فارغ التحصیلان نمی باشد بدین جهت ایجاد شرایط مشابه محیط کار واقعی به هنگام انجام آزمون و ارزیابی ضروری است [۱۸]. با توجه به اینکه آزمون های چند گزینه ای به تنهایی نمی توانند وضعیت دانشجو را بصورت کامل و همه جانبه مورد سنجش قرار دهند طی دهه های اخیر استفاده از آزمون های مبتنی بر عملکرد مورد توجه قرار گرفته، تحلیل و ارزیابی این آزمون ها ضروری است زیرا اطمینان از میزان صحت عملکرد آزمون هم برای شرکت کنندگان و هم برگزار کنندگان ضروری است [۱۹].

یکی از روشهای ارزشیابی مبتنی بر عملکرد، آزمون عینی ساختارمند بالینی می‌باشد که از دهه هفتاد میلادی ابداع و به کار گرفته شده و بطور روزافزون مورد توجه قرار گرفته است. در سالهای اخیر استفاده از این روش در دانشگاههای علوم پزشکی کشورمان افزایش چشمگیری داشته است [۱۹]. آسکی یک روش ارزشیابی است



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

که مراحل بالینی مانند شرح حال، معاینه مهارت های ساده، کاربرد نتایج آزمایشگاهی، مشکلات موجود در کنترل و هدایت بیمار، ارتباط و تعامل بین فردی، و نگرش را مورد ارزیابی قرار می دهد و با چک لیستی عملکرد دانشجویان را در ایستگاه های مختلف اندازه گیری می کند[۱۵]. این در حالیست که غالبا" آزمون های برگزار شده مورد بررسی و تحلیل قرار نمی گیرد و تنها شاخص پایایی که اطلاعاتی در مورد کل آزمون را ارائه می کند مورد توجه قرار می‌گیرد. نکته مهم اینجاست که برای تحلیل آزمون عینی ساختارمند بالینی تنها استفاده از یک شاخص کافی نیست و شاخص های دیگری وجود دارد که اطلاعات مناسبی در مورد کل امتحان و هر کدام از ایستگاه ها به دست می دهد؛ به منظور تحلیل کیفیت آزمون ضروری است که شاخص های سایکومتریک نیز مورد توجه و اندازه گیری قرار گیرد [۲۰،۲۱]. با توجه به اینکه آزمون های ساختارمند عینی در دانشگاه علوم پزشکی البرز از سال ۱۳۹۷ تاکنون به طور پیوسته انجام شده است این پژوهش با هدف تحلیل کیفیت این آزمون انجام شد.

## روش

این مطالعه از نوع کمی و توصیفی تحلیلی است. جامعه پژوهش شامل ۶۴ شرکت کننده از دانشجویان سال آخر رشته پزشکی که به روش سرشماری درمرکز مهارتهای بالینی دانشگاه علوم پزشکی البرز طی سه دوره (۳۱، ۳۲، ۳۳) آزمون آسکی دوره کارورزی در سال ۱۴۰۰ وارد مطالعه شدند. آزمون آسکی به عنوان یک آزمون عملی بالینی شبیه دیگر آزمون ها نظیر سوالات شفاهی، چند گزینه ای نبوده بلکه یک چهارچوب سازمان دهی شده شامل چند ایستگاه است. فضای آزمون شامل اتاق هایی است که اصطلاحاً ایستگاه نامیده می شود. در هر ایستگاه یک مهارت کلی که خود می تواند شامل چندین مهارت باشد مورد سنجش قرار می گیرد. برای هر ایستگاه سناریوی متفاوت نوشته شد که بر اساس سناریوی مورد نظر تجهیزات و وسایل مورد نظر هر ایستگاه تهیه شد و به طور مناسب چیده شد. برای هر ایستگاه دستورالعملی نوشته شد و بر روی در نصب گردید که دانشجو باید قبل از ورود به ایستگاه آن را مطالعه نماید. برای هر ایستگاه مدت زمان لازم برای انجام هر عملکرد ۵ دقیقه در نظر گرفته شد و طی آن دانشجو تمامی کارهای خواسته شده را به تنهایی انجام داد. هر ایستگاه یک آزمونگر تخصصی مستقر در این مکان را داشت که به صورت عینی نحوه انجام مهارت ها توسط دانشجو را بر اساس چک لیست تهیه شده سنجیده شد.

سی و یکمین دوره آزمون صلاحیت بالینی که از ۱۵ ایستگاه و۲۷ شرکت کننده تشکیل شده بود. ایستگاههای ۲ و ۳و۵ و ۱۰و ۱۵( پاتولوژی، رادیولوژی، داخلی، اطفال، جراحی) بدون آزمون گرمستقیم و نمره گلوبال طراحی شده بود از آنجا که نحوه نمره دهی آنها مانند سوالات تشریحی بود این پنج ایستگاه و دو ایستگاه استراحت (۴و۱۲) از مطالعه خارج شدند. و به این ترتیب اطلاعات هشت ایستگاه ( اورولوژی، زنان و زایمان، داخلی قلب، ارتوپدی، پوست، اورژانس، پزشکی اجتماعی، داخلی ریه)وارد مطالعه شدند.

سی و دومین دوره آزمون صلاحیت بالینی که از ۱۵ ایستگاه و ۲۳ شرکت کننده تشکیل شده بود ایستگاه های۲و۳و۵و۶ (پاتولوژی، عفونی، اورولوژی، زنان) بدون آزمونگر مستقیم و نمره گلوبال طراحی شده بود. از آنجا که نحوه نمره دهی آنها مانند سوالات تشریحی بود این چهار ایستگاه و دو ایستگاه استراحت(۴و۱۲)از مطالعه خارج شدند. به این ترتیب اطلاعات ۹ ایستگاه (داخلی کلیه، طب اورژانس، داخلی ریه، روانپزشکی، اطفال،رادیولوژی، پوست، قلب، جراحی) وارد مطالعه شدند. سی و سومین دوره آزمون صلاحیت بالینی که از ۱۴ ایستگاه و ۱۴ نفر شرکت کننده تشکیل شده بود ایستگاه های ۱و۲و۳ (داخلی، روانپزشکی، پوست) بدون آزمونگر مستقیم و نمره گلوبال طراحی شده بود. از آنجا که نحوه نمره دهی آنها مانند سوالات تشریحی بود این سه ایستگاه و دو ایستگاه استراحت (۴و۱۴) از مطالعه خارج شدند. و به این ترتیب اطلاعات۹ ایستگاه (ارتوپدی، طب اورژانس، زنان، پزشکی اجتماعی، کودکان، رادیولوژی، داخلی غدد، اورولوژی، جراحی) وارد مطالعه شدند. در این مطالعه دانشجویان سال آخر که تمایل به شرکت در آزمون داشتند و حداقل شش ماه از شروع دوره کارورزی آنان گذشته بود شرکت نمودند. در این آزمون آسکی دانشجویان ثابت دانشگاه علوم پزشکی البرز وارد مطالعه شدند. دانشجویان مهمان از مطالعه حذف شدند. قبل از شروع آزمون برای تمام آزمونگرهای تخصصی در مورد نحوه دادن نمره به آنها توضیح داده شد مبنی بر اینکه برای هر ایستگاه باید دو نمره به دانشجو داده شود. ۱- نمره چک لیست که باید براساس چک لیست وزارتخانه



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

به دانشجو از ۲۰ نمره داده شود .۲- نمره گلوبال در هر ایستگاه بر اساس ارزیابی کلی آزمونگر نسبت به عملکرد دانشجو به صورت یک طیف لیکرت ۵ تایی (رد، مرزی، قابل قبول، خوب، عالی) به دانشجو داده شود و نمره لیکرت واقعاً باید منعکس کننده عملکرد کلی دانشجو باشد و آزمونگر نباید نمره چک لیستی که دانشجو در ایستگاه کسب کرده را به نمره لیکرت تبدیل کند. همچنین به دانشجویان نیز قبل از شروع آزمون در مورد اهداف، نحوه برگزاری، زمان لازم برای انجام عملکرد ها و ترتیب حضور در ایستگاه ها توضیح داده شد و از آنان رضایت نامه کتبی آگاهانه اخذ گردید. برای انجام آزمون در هر یک از ایستگاه ها پنج دقیقه فرصت وجود داشت تا به انجام وظیفه بپردازند. هر ایستگاه دارای یک آزمونگر تخصصی مستقل بود . نمره گلوبال صرفاً برای تعیین حدنصاب نمره آزمون و تعیین شاخص های سایکومتریک مورد استفاده قرار گرفت که حد نصاب نمره آزمون و شاخص های اندازه گیری در این پژوهش توضیح داده می شود.

## تعیین حدنصاب نمره

روش رگرسیون مرزی[1] یکی از تکنیک های رایج برای تعیین نمره قبولی در آزمون های مبتنی بر عملکرد می باشد. [15] در یک مطالعه مقایسه چهار روش نمره ثابت، آنگوف، رگرسیون مرزی، کوهن مشخص کرد که روش رگرسیون مرزی بیشترین روایی و پایایی برای بررسی حد نصاب آزمون های معیار-محور[2] دارا می باشد. [23] برای هر ایستگاه یک معادله رگرسیون طراحی شد. نمرات چک لیست به عنوان متغیر وابسته و نمرات گلوبال به عنوان متغیر مستقل وارد معادله رگرسیون شدند. سپس با قرار دادن نمره گلوبال مربوط به دانشجوی مرزی در معادله (۲ در مقیاس لیکرت)، حد نصاب ایستگاه بدست می آید و حد نصاب کل آزمون با پیش گرفتن رویکرد جبرانی[2] از جمع حد نصاب ایستگاه ها به دست می آید[15].

برای تخمین میزان خطای روش تعیین حدنصاب در کل آزمون از ریشه میانگین مربعات خطا[3] طبق فرمول زیر استفاده خواهد شد. در این فرمول $M$ تعداد ایستگاه ها، $n$ تعداد دانشجویان و $S_{regr,\,i}$ انحراف معیار معادله رگرسیون برای ایستگاه $i$ هستند . همچنین $Mean_{G,i}$ و $SD_{G,i}$, $I$ به ترتیب میانگین و انحراف معیار نمرات گلوبال برای ایستگاه $i$ هستند. $G_0$ نیز نمره گلوبالی است که برای دانشجوی مرزی در نظر گرفته شده و در تمام ایستگاه های این مطالعه یکسان و معادل ۲خواهد بود[15].

$$\boldsymbol{RMSE_{OSCE}} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{m^2}}}\cdot\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{i=1}^{m}\{S^2_{reger,i}\left(1+\frac{(G_0-Mean_{G,i})}{[(n-1)/n]\,.\,SD^2_{G,i}}\right)\}$$

## شاخص های سایکومتریک کل آزمون

ابتدا شاخص های میانگین و انحراف معیار برای کل آزمون محاسبه شد و سپس میزان ردی کل بر حسب درصد در کل آزمون محاسبه شد.

## شاخص های سایکومتریک ایستگاهها

ابتدا میانگین، انحراف معیارو میزان ردی را در هر کدام از ایستگاه ها محاسبه شد. از معادله رگرسیونی که برای هر ایستگاه تعیین شده بود شاخص $R^2$ محاسبه شد. ضریب تمیز و ضریب دشواری در هر ایستگاه با توجه به روش ویتنی و سیبرز برای سوالات تشریحی محاسبه شد.

## آنالیز آماری

تجزیه و تحلیل داده ها با استفاده از نرم افزار SPSS 26 و Excel 16 انجام شد. آنالیز رگرسیون خطی با در نظر گرفتن نمره چک لیست(متغیر وابسته) و نمره گلوبال (متغیر مستقل) انجام شد که براساس معادله رگرسیون علاوه بر پیش بینی دانشجوی مرزی $R^2$ و RMSE محاسبه شد.

## نتایج

تعداد کل دانشجویان شرکت کننده در مطالعه ۶۴ نفر بود که در آزمون ۳۱ تعداد ۲۷ نفر و در آزمون ۳۲ تعداد ۲۳ نفر و در آزمون ۳۳ تعداد ۱۴ نفر شرکت کرده بودند. در سی و یکمین آزمون صلاحیت بالینی ۶۶/۷ درصد از دانشجویان مرد و ۳۳/۳ درصد زن و همچنین در سی و دومین آزمون صلاحیت بالینی ۴۷/۸ درصد از دانشجویان مرد و ۵۲/۲ درصد زن و سی و سومین آزمون صلاحیت بالینی ۶۴/۳ درصد از دانشجویان مرد و ۳۵/۷ درصد زن بودند.

[1] Borderline Regression Method
[2] compensatory
[3] Root Mean Square Error



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]
[ DOI: 10.61186/aums.12.4.485 ]

جدول ۱: شاخصهای توصیفی (مینیمم، ماکزیمم، میانگین و انحراف معیار، ضریب تعیین، ریشه میانگین مربعات خطا و حد نصاب قبولی) برای هر ایستگاه در آزمون صلاحیت بالینی ۳۱، ۳۲ و۳۳)

| | نام ایستگاه | حداقل | حداکثر | میانگین | انحراف معیار | ضریب تعیین $R^2$ | RMSE | حد نصاب قبولی | فراوانی ردی(درصد) |
|---|---|---|---|---|---|---|---|---|---|
| آزمون سی و یکم | اورولوژی | ۱۱ | ۱۸ | ۱۶/۸۹ | ۱/۴۸ | **۰/۸۵** | ۰/۵۹ | ۱۲/۳ | ۱(۳/۷) |
| | زنان | ۷ | ۱۹ | ۱۴/۳۰ | ۲/۸۳ | ۰/۶۳ | ۱/۷۵ | ۱۲/۰۳ | ۲(۴/۷) |
| | داخلی-قلب | ۱۲ | ۱۹ | ۱۵/۷۸ | ۲/۰۳ | ۰/۶۳ | ۱/۲۶ | ۱۲/۶۳ | ۰(۰) |
| | ارتوپدی | ۱۲ | ۲۰ | ۱۸/۲۶ | ۱/۸۱ | ۰/۵۶ | ۱/۲۲ | ۱۳/۷۸ | ۱(۳/۷) |
| | پوست | ۹/۵ | ۲۰ | ۱۷/۴۴ | ۳/۲۶ | **۰/۴۵** | ۱/۵۵ | ۱۱/۵ | ۴(۱۴/۸) |
| | اورژانس | ۱۴ | ۲۰ | ۱۸/۰۴ | ۱/۶۱ | ۰/۶۴ | ۰/۹۸ | ۱۳/۹ | ۰ |
| | پزشکی اجتماعی | ۱۵ | ۲۰ | ۱۸/۶۴ | ۱/۷۰ | ۰/۸۱ | ۰/۷۵ | ۱۳/۵ | ۰ |
| | داخلی-ریه | ۱۲ | ۱۹ | ۱۵/۴۳ | ۱/۹۱ | ۰/۵۲ | ۱/۳۴ | ۱۲/۹ | ۰ |
| | کل آزمون | ۱۳/۷۵ | ۱۷/۹۱ | ۱۶/۸۵ | ۱/۰۴ | ۰/۷۴ | ۰/۵۵ | ۱۳ | ۰ |
| آزمون سی و دوم | داخلی-کلیه | ۵ | ۱۸ | ۱۴/۸۷ | ۳/۳۶ | **۰/۸۷** | ۱/۲۲ | ۱۰/۲۸ | ۲(۸/۶) |
| | طب اورژانس | ۱۵ | ۲۰ | ۱۸/۰۴ | ۱/۴۰ | **۰/۴۴** | ۱/۰۷ | ۱۴/۶۰ | ۰ |
| | داخلی-ریه | ۱۵ | ۲۰ | ۱۷/۶۵ | ۱/۴۳ | ۰/۵۴ | ۰/۹۹ | ۱۲/۹۹ | ۰ |
| | روانپزشکی | ۱۰ | ۱۹ | ۱۵/۵۲ | ۲/۰۹ | ۰/۷۶ | ۱/۰۵ | ۱۲/۹۱ | ۱(۴/۳) |
| | اطفال | ۱۵ | ۱۹ | ۱۷/۲۲ | ۱/۱۳ | ۰/۶۴ | ۰/۶۹ | ۱۴/۲۵ | ۰ |
| | رادیولوژی | ۱۲ | ۱۸/۵ | ۱۲/۹۸ | ۱/۵۲ | ۰/۵۲ | ۱/۰۷ | ۱۱/۲۶ | ۰ |
| | پوست | ۹ | ۱۸ | ۱۳/۵۲ | ۲/۷۶ | **۰/۳۸** | ۲/۲۲ | ۱۲/۷۲ | ۹(۳۳/۳) |
| | قلب | ۶ | ۱۶/۵ | ۱۳/۲۸ | ۲/۸۴ | ۰/۷۴ | ۱/۴۷ | ۱۱/۳۴ | ۳(۱۳) |
| | جراحی | ۱۲ | ۲۰ | ۱۷/۳۳ | ۲/۶۵ | ۰/۸۶ | ۱/۰۱ | ۱۳/۴۱ | ۱(۴/۳) |
| | کل آزمون | ۱۳/۶۱ | ۱۷/۰۶ | ۱۵/۶۰ | ۰/۸۶ | ۰/۵۷ | ۰/۵۷ | ۱۲/۵۸ | ۱(۴/۳) |
| آزمون سی ودوم | ارتوپدی | ۱۲ | ۱۸ | ۱۴/۳۲ | ۱/۸۷ | **۰/۳۹** | ۱/۵۳ | ۱۲/۲۶ | ۳(۲۱/۴) |
| | طب اورژانس | ۷ | ۲۰ | ۱۳/۰۴ | ۳/۶۹ | ۰/۷۰ | ۲/۱۱ | ۱۱/۹۸ | ۵(۳۵/۷) |
| | زنان | ۹ | ۲۰ | ۱۶/۹۳ | ۳/۴۵ | ۰/۷۲ | ۱/۸۸ | ۱۰/۷۳ | ۱(۷/۱) |
| | پزشکی اجتماعی | ۱۱ | ۲۰ | ۱۵/۵۰ | ۲/۳۴ | ۰/۴۲ | ۱/۸۵ | ۱۳/۵۶ | ۲(۱۴/۲) |
| | کودکان | ۱۳ | ۱۸ | ۱۵/۳۶ | ۱/۲۸ | ۰/۴۱ | ۱/۰۲ | ۱۲/۵۰ | ۰ |
| | رادیولوژی | ۱۲ | ۱۸ | ۱۵/۲۹ | ۲/۳۱ | ۰/۵۴ | ۱/۶۳ | ۱۰/۲۹ | ۰ |
| | داخلی-غدد | ۶ | ۱۸ | ۱۰/۸۹ | ۳/۱۰ | ۰/۵۶ | ۲/۱۴ | ۹/۲۶ | ۵(۳۵/۷) |
| | اورولوژی | ۱۵ | ۲۰ | ۱۷/۴۳ | ۱/۴۰ | ۰/۷۹ | ۰/۶۷ | ۱۲/۷۶ | ۰ |
| | جراحی | ۱۳ | ۲۰ | ۱۹/۰۰ | ۱/۸۸ | **۰/۹۵** | ۰/۴۳ | ۱۴/۶۷ | ۱(۷/۱) |
| | کل آزمون | ۱۲/۲۲ | ۱۷/۳۱ | ۱۵/۳۰ | ۱/۲۹ | ۰/۶۸ | ۰/۷۷ | ۱۱/۹۳ | ۰ |

نمره هر ایستگاه و نمره کل آزمون از ۲۰ است.



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

همانطور که در جدول ۱ مشاهده می شود بالاترین حد نصاب در آزمون سی و یکم مربوط به ایستگاه اورژانس(۱۳/۹)و کمترین حد نصاب مربوط به ایستگاه پوست(۱۱/۵) بود. میانگین و انحراف معیار نمرات در کل آزمون به ترتیب ۱۶/۸۵ و ۱/۰۴ بدست آمد. با محاسبه نمره کل تمام دانشجویان و با توجه به نمره حد نصاب کل(۱۳) مشخص شد همه دانشجویان در این آزمون نمره قبولی را آورده اند و مردود نداشتیم. با توجه به اینکه میزان بالای ضریب تعیین (بالای ۰/۵)حاکی از پیشگویی قوی و مقدار RMSE کمتر نشان از میزان خطای کمتر برای پیشگویی است. بیشترین ضریب تعیین ۰/۸۵ و کمترین RMSE ۰/۵۹مربوط به ایستگاه اورولوژی می‌باشد. در آزمون سی و دوم بالاترین حد نصاب مربوط به ایستگاه طب اورژانس (۱۶/۴۰)و کمترین حد نصاب مربوط به ایستگاه داخلی کلیه(۱۰/۲۸) بود. میانگین نمرات در کل آزمون ۱۵/۶۰ با انحراف معیار کل ۰/۸۶ بدست آمد. با محاسبه نمره کل تمام دانشجویان و با توجه به نمره حد نصاب کل ۱۲/۵۸مشخص شد میزان ردی ۴/۳درصد یعنی یک مورد بود. بیشترین ضریب تعیین $R^2$ مربوط به ایستگاه داخلی کلیه ۰/۸۸ و کمترین RMSE مربوط به ایستگاه اطفال ۰/۶۹ است.

در سی و سومین آزمون بالاترین حد نصاب مربوط به ایستگاه جراحی(۱۴/۶۷)و کمترین حد نصاب مربوط به ایستگاه داخلی غدد(۹/۲۶) بود. میانگین نمرات در کل آزمون ۱۵/۳۰ با انحراف معیارکل ۱/۲۹ بدست آمد. بامحاسبه نمره کل تمام دانشجویان و با توجه به نمره حد نصاب کل که ۱۱/۹۳مشخص شد همه دانشجویان در این آزمون نمره قبولی را آورده اند و مردودی نداشتیم و میزان ردی کل آزمون صفر درصد بود. بیشترین ضریب تعیین مربوط به ایستگاه جراحی ۰/۹۵ و کمترین RMSE مربوط به جراحی۰/۴۳ بود.

## سنجش میزان ضریب دشواری و ضریب تمیز در آزمون صلاحیت بالینی(۳۱،۳۲،۳۳)

از معیارهای تحلیل سوالات آزمون ضریب دشواری و تمیز می‌باشد. ضریب دشواری عددی بین صفر و یک است و اگر این عدد به یک نزدیک باشد یعنی درصد بالایی از دانشجویان به سوال پاسخ صحیح داده اند و اگر نزدیک صفر باشد نشان دهنده آن است که درصد بالایی به این سوال پاسخ غلط داده اند. در صورتی که ضریب تمیز، قدرت سوال را در تمایز بین گروه قوی و گروه ضعیف آزمون شوندگان را مشخص می کند. این شاخص نیز عددی بین صفر و یک است و هر چقدر این عدد بزرگتر باشد متمایز کردن افراد قوی و ضعیف بهتر است. برای محاسبه ضریب دشواری و ضریب تمیز در هر ایستگاه از آزمون ویتنی و سیبرز استفاده شد. طبق جدول شماره ۲- در سی و یکمین آزمون سخت ترین و آسانترین ایستگاه ها بر اساس ضریب دشواری به ترتیب عبارتند از ایستگاه زنان(۰/۳۷) و ایستگاه پزشکی اجتماعی(۰/۴۷) بودند. ضریب تمیز تمامی ایستگاه ها کمتر یا مساوی۰/۰۶ بود که مقدار بسیار پایینی است. در سی و دومین آزمون نشان داده شد که سخت ترین و آسانترین ایستگاه ها بر اساس ضریب دشواری به ترتیب عبارتند از ایستگاه رادیولوژی (۰/۲۰) و ایستگاه طب اورژانس (۰/۲۸) بود. ضریب تمیز تمامی ایستگاه ها بین ۰/۱۵ الی ۰/۲ بود که مقدار نسبتا پایینی است. در سی و سومین آزمون سخت ترین و آسانترین ایستگاه ها بر اساس ضریب دشواری به ترتیب ایستگاه داخلی غدد(۰/۲۷) و ایستگاه جراحی (۰/۴۸) بود. ضریب تمیز تمامی ایستگاه ها بجز اورولوژی (۰/۲۳)کمتر یا مساوی ۰/۰۶ بود که مقدار بسیار پایینی است. (طبق جدول۲)

ضریب دشواری بین ۰/۷-۰/۳ مناسب و در هر سه دوره آزمون ضرایب دشواری کلیه ایستگاه ها مقدار قابل قبولی داشتند. ضریب تمیز (۰/۳-۰)کم و متاسفانه در هر سه آزمون ضرایب تمیز در کلیه ایستگاه ها مقدار قابل قبولی به دست نیامده است.



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

جدول ۲-متریک های مربوط به هر یک از ایستگاه های OSCEدر آزمون های صلاحیت بالینی (۳۱، ۳۲، ۳۳)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| آزمون ۳۱ | نام ایستگاه | اورولوژی | زنان | داخلی-قلب | ارتوپدی | پوست | اورژانس | پزشکی اجتماعی | داخلی ریه | |
| | ضریب دشواری | | | | | | | | | |
| | ضریب تمیز | | | | | | ۰/۰۳ | | | |
| آزمون ۳۲ | نام ایستگاه | داخلی-کلیه | طب اورژانس | داخلی-ریه | روانپزشکی | اطفال | رادیولوژی | پوست | قلب | جراحی |
| | ضریب دشواری | | | | | | | | | |
| | ضریب تمیز | | | | | | | | | |
| آزمون ۳۳ | نام ایستگاه | ارتوپدی | طب اورژانس | زنان | پزشکی اجتماعی | کودکان | رادیولوژی | داخلی-غدد | اورولوژی | جراحی |
| | ضریب دشواری | | | | | | | | | |
| | ضریب تمیز | | | | | | | | | |

## سنجش میزان ارتباط بین نمره کسب شده در هر ایستگاه با معدل نمرات ای ستگاه ها با ا ستفاده از ضریب همب ستگی پیرسون آزمون صلاحیت بالینی بالینی(۳۳،۳۲،۳۱)

ضریب همبستگی شاخصه ای است برای سنجش رابطه خطی بین دو ویژگی کمی، بطوری که هر چقدر قدر مطلق مقدار آن به یک نزدیک باشد شدت رابطه بین دو ویژگی بیشتر است. بدین معنی که می توان از یکی از ویژگی ها برای برآورد ویژگی دیگر ا ستفاده نمود. برای سنجش رابطه نمره هر ای ستگاه با معدل نمرات ایستگاه ها از ضریب همبستگی پیرسون استفاده کردیم.

نتایج به دست آمده نشان داد که در سی و یکمین آزمون صلاحیت بالینی دان شجویان پز شکی بین نمرات ای ستگاه پز شکی اجتماعی با میانگین نمره کل همبستگی بالا، بین ایستگاه‌های زنان و ارتوپدی و پوست با میانگین نمره کل همبستگی متوسط و سایر ایستگاه‌ها همبستگی پایینی داشتند.

در سی و دومین آزمون بین نمرات ایستگاه داخلی-کلیه وپوست با میانگین نمره همب ستگی متو سط و سایر ای ستگاه‌های همب ستگی پایینی داشتند.

در سی و سومین آزمون بین نمره ای ستگاه زنان و رادیولوژی با متوسط نمره کل رابطه قوی و سایر ایستگاه ها رابطه ضعیفی وجود داشت.(طبق جدول۳)

جدول ۳- میزان ارتباط بین نمره کسب شده در هر ایستگاه با معدل نمرات ایستگاه با استفاده از ضریب همبستگی پیرسون آزمون صلاحیت بالینی (۳۱، ۳۲، ۳۳)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| آزمون ۳۱ | نام ایستگاه | اورولوژی | زنان | داخلی-قلب | ارتوپدی | پوست | اورژانس | پزشکی اجتماعی | داخلی ریه | |
| | ضریب دشواری | ۰/۴۲ | ۰/۳۶ | ۰/۳۹ | ۰/۴۶ | ۰/۴۳ | ۰/۴۵ | ۰/۴۶ | ۰/۳۹ | |
| | ضریب تمیز | ۰/۰۲ | ۰/۰۵ | ۰/۰۴ | ۰/۰۳ | ۰/۰۶ | ۰/۰۳ | ۰/۰۳ | ۰/۰۴ | |
| آزمون ۳۲ | نام ایستگاه | داخلی-کلیه | طب اورژانس | داخلی-ریه | روانپزشکی | اطفال | رادیولوژی | پوست | قلب | جراحی |
| | ضریب دشواری | ۰/۲۳ | ۰/۲۸ | ۰/۲۷ | ۰/۲۴ | ۰/۲۶ | ۰/۲۰ | ۰/۲۲ | ۰/۲۲ | ۰/۲۸ |
| | ضریب تمیز | ۰/۱۹ | ۰/۲۰ | ۰/۲۰ | ۰/۱۹ | /۱۹۰ | ۰/۱۵ | ۰/۱۷ | ۰/۱۷ | ۰/۲۱ |
| آزمون ۳۳ | نام ایستگاه | ارتوپدی | طب اورژانس | زنان | پزشکی اجتماعی | کودکان | رادیولوژی | داخلی-غدد | اورولوژی | جراحی |
| | ضریب دشواری | ۰/۳۶ | ۰/۳۳ | ۰/۴۲ | ۰/۳۹ | ۰/۳۸ | ۰/۳۸ | ۰/۲۷ | ۰/۴۴ | ۰/۴۸ |
| | ضریب تمیز | ۰/۰۳ | ۰/۰۶ | ۰/۰۵ | ۰/۰۴ | ۰/۰۲ | ۰/۰۴ | ۰/۰۵ | ۰/۲۳ | ۰/۰۱ |

## بحث

در این مطالعه شاخص های سایکومتریک ( ضریب د شواری، ضریب تمیز، ضریب تعیین و مقدار RMSE آزمون OSCE مقطع کارورزی دانشگاه علوم پز شکی البرز در سطح تک تک ایستگاه ها و در سطح کل ایستگاه ها مورد بررسی قرار گرفت.



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

در این مطالعه میزان ردی با توجه به حد نصاب نمره با روش رگرسیون مرزی ۱/۵ درصد بود که در مقایسه با میزان ردی آزمون سال ۸۸ دانشگاه تهران که ۴/۸ درصد محاسبه شده بود [۲۳] و همچنین در مقایسه با میزان ردی در مطالعه Kossay Elabd که ۲/۷ درصد می‌باشد، میزان ردی در مطالعه ما پایین تر است. اگرچه تفاوت خیلی زیاد نیست[۲۵].

مقدار RMSE کل در این مطالعه بین ۰/۵۵ الی ۰/۷۷ بود که در مقایسه با مقدار آن در آزمون سال ۱۳۸۸ دانشگاه تهران که ۰/۵۵ بود و سال ۱۳۹۱ دانشگاه تهران ۰/۴۵ به دست آمده بود میزان آن فقط در آزمون سی و سوم بالاتر می باشد [۲۳]. در مطالعه ای که Homer و همکاران برای بررسی میزان خطای روش رگرسیون مرزی در حجم نمونه های متفاوت انجام دادند، دریافتند که میزان خطا در OSCE هایی که با شرکت تعداد زیاد دانشجویان (حدود ۲۵۰تا ۳۰۰) برگزار می شوند، چندان زیاد نیست. این نویسندگان میزان RMSE برای یک آزمون ۲۰ ایستگاهه را حدود ۰/۳ برآورد کردند و ذکر نمودند که با کاهش تعداد دانشجویان این مقدار افزایش می یابد [۲۲]. با توجه به تعداد پایین بودن تعداد دانشجویان مقدار ۰/۷۷ چندان دور از انتظار نیست.

شاخص $R^2$ (ضریب تعیین)[۴] میزان تغییر در نمره چک لیست(متغیر وابسته) را نسبت به تغییراتی که در نمره گلوبال (متغیر مستقل) ایجاد شده نشان می دهد و با توجه به اینکه قبل از آزمون به آزمونگرها توضیح داده شده که در ایستگاه نمره چک لیست را برای دادن نمره گلوبال ترجمه نکرده باشند و بطور مستقل به عملکرد دانشجو امتیاز دهند در غیر اینصورت مقدار شاخص ضریب تعیین به صورت غیر طبیعی بالا خواهد بود. در مطالعه ما بالاترین مقدار ضریب همبستگی پیرسون در سی و سومین آزمون در ایستگاه جراحی ۰/۹۵و کمترین آن مربوط به آزمون سی و دوم و ایستگاه پوست ۰/۳۸ بود. از آنجایی که مقادیر بین ۰/۵ الی ۰/۷۵ نشان از رابطه قابل قبول بین دو نمره دارد. مقادیر پایین نشان از عدم هماهنگی و مقادیر بالا نیز ممکن است حاکی از عدم توجه آزمونگر به مستقل بودن دو نمره و یا وجود مشکلی در طراحی چک لیست می‌باشد. ضریب تعیین در مطالعه ما در مقایسه با مطالعه سونگ پارک (ref) که ضریب تعیین بین ۰/۲۸ الی ۰/۶۰ مقادیر بالاتری و همچنین مطالعه چین کوهونگ و همکارانش که بین ۰/۰۱ و ۰/۴۱ بالاتر به دست آمده است که بنظر می رسد آزمونگرها به استقلال نمره چک لیست و نمره گلوبال توجه نکرده اند[۲۴ و ۲۶].

مقدار ضریب دشواری در این مطالعه بین ۰/۲۰الی ۰/۴۸ بود که مقادیر بین ۰/۳ الی ۰/۷ قابل قبول می باشد. در بیش از نیمی از ایستگاه ها مقدار آن از ۰/۳ بیشتر بوده است. در مقایسه با مطالعه مرتاض میزان دشواری مطالعه ما بالا تر بوده یا اینکه بنیه علمی دانشجویان پایین است که با توجه به تیپ ۳ بودن دانشگاه و داشتن رتبه های بالاتر در کنکور ورودی نتیجه به دست آمده چندان دور از ذهن نیست. در رابطه با ضریب تمیز بین ۰/۲۳-۰/۰۲ بود در مقایسه با مطالعه مرتاض که ضریب تمیز در آن ۰/۱۲ الی ۰/۴۴ بود، نشان دهنده آن است که آزمون نتیجه رضایت بخشی را نداشت و آزمون امکان تفکیک بین دانشجویان قوی و ضعیف را نداشت.

ضریب همبستگی بین نمره هر ایستگاه با نمره کل نشان دهنده میزان توافق نمره دانشجو در هر ایستگاه با نمره نهایی می باشد. در مطالعه ما در بیش از نیمی از ایستگاه ها مقدار ضریب همبستگی از۰/۵ کمتر بود ولی فقط ۱۲ درصد ایستگاه ها دارای ضریب همبستگی کمتر از ۰/۲ بودند در مقایسه با مطالعه چین کوهونگ و همکارانش که میزان همبستگی آیتم کل با نمره هر ایستگاه بین ۰/۳۲ و ۰/۶۴ بود، ضرایب همبستگی پایین نشان دهنده آن است که دانشجو در یک ایستگاه خاص عملکرد متفاوتی را در مقایسه با کل آزمون داشته است که در مطالعه ما بنظر می رسد با توجه به ضرایب دشواری و تمیز سوالات نیاز به بازنگری دارد[۲۴].

از نقاط قوت این مطالعه می توان به این موضوع اشاره کرد که استفاده از دو ممتحن برای ارزیابی هر ایستگاه موجب ارزیابی دقیق‌تری می شود. بطوری که یک ممتحن ارزیابی عملکرد دانشجو را براساس چک لیست و دیگری عملکرد کلی را با استفاده از یک سیستم امتیازدهی گلوبال انجام می دهد که البته نیاز به نیروی متخصص را افزایش می دهد. از محدودیت های این مطالعه می توان به گذشته نگر بودن آن، تعداد کم دانشجویان در هر آزمون اشاره کرد. همچنین این احتمال که ارزیابی نهایی ممتحن از عملکرد دانشجو می‌تواند تحت تأثیر امتیاز آنها در چک لیست قرار گیرد که

[۴] coefficient of determination



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

برای رفع این م حدود یت می توان از کلاس های بازآموزی و توجیهی پیش از برگزاری امتحان استفاده نمود.

## نتیجه گیری

شـــاخص های ســـایکومتر یک در مورد کیفیت آزمون های صـلاحیت بالینی مورد اسـتفاده قرار می گیرد. نتایج به دسـت آمده نشان می دهد که ضریب تمیز بسیار پایین تر از مقدار قابل قبول بود و ســـوالات برای آزمون های بعدی نیاز به بازنگری مجدد دارد و نیازمند محا سبه مجدد شاخص های سایکومتریک برای ر سیدن به مقدار مطلوب را دارد. در رابطه با سایر شاخص ها مقادیر به د ست آمده تقریبا مناسب و قابل قبول بودند.

## پیشنهادات

باتوجه به اهمیت ویژگی سایکومتریک در آزمون های آموزش پزشـــکی، مخصـــوصـــا"در آزمون های مبتنی بر عملکرد که دارای حســـاســـیت بالایی هســـتند به کارگیری تحلیل آزمون ها با توجه به شـــاخص های ســـایکومتریک به اســـاتید و متولیان آموزش عالی و سیا ستگذاران و برنامه ریزان آزمون ها پی شنهاد می گردد تا بتوانند نقاط ضعف و نقاط قوت این آزمون ها را بررسی کنند.



[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

## References


1. Mlek M, Amico M D, Naseem A, Waddington D, Hamalian A/Nursing Students learning Experiences in Clinical Settings:Stress,Anxiety and Coping/ Ma Thesis/ Canada:Concordia University/2011/
2. Craven R/ Fundamentals of nursing: Human Health and Function/ 4th ed/ Philadelphia: Wolters Kluwer/ 2003/ PP: 137–162, 362-371/
3. Saarikoski M, Kaila P, Lambrinou E, Pérez Cañaveras RM, Tichelaar E, Tomietto M, et al/ Students' experiences of cooperation with nurse teacher during their clinical placements: an empirical study in a Western European context/ Nurse Educ Pract/ 2013 Mar/13(2)/pp:78-82/
4. Mongwe RN, Van der Wall DM/ Students Nurses experiences of the clinical field in the Limpopo province as leraning Field: A Phenomenological study/ PhD Thesis/ University of South Africa/2007/
5. Pap I, Markkanen M, Von B/ Clinical environment as a learning environment:Student nurse perceptions concerning clinical learning experiences/ Nurse Educ today/ May2003/23(4)/ PP:262-268/
6. Hassani p, Cheraghi F, Yaghmaei F/ self-efficacy and elfregulted learning in clinical performance of nursing students: A qualitative research/ Iranian Jornal of medical education/2008 spr-sum/ 8(1)/PP:33-41/
7. Frankel A/ Nurses' learning styles: promoting better integration of theory into practice/ Nurs Times/ 2009 Jan/105(2)/pp:24-7/
8. Hartigan Rogers JA// Nursing Graduate perception of clinical placement/ International journal of nursing education scholarship 2007;4(2):1-14/
9. Chesser-Smyth PA/ The lived experiences of general student nurses on their first clinical placement: A phenomenological study/ *Nurse Educ Pract* 2005;5(6):320-7/
10. Andrews GJ, Brodie DA, Andrews JP, Hillan E, Gail Thomas B, Wong J, et al/ Professional roles and communications in clinical placements: A qualitative study of nursing students' perceptions and some models for practice/ Int J Nurs Stud 2006;43(7):861-74/
11. Dickson C, Walker J, Bourgeois S/ Facilitating undergraduate nurses clinical practicum: the lived experience of clinical facilitators/ Nurse Educ Today 2006;26(5):416-22/
12. Peyrovi H/ Student nurses experiences of clinical placement/ A Dissertation presented in PhD of nursing/Iran University of Medical Science/ 2005/
13. Ebrahimi H, Sanagoo A, Jouybari L/ Experiences of Nursing Students about stress in clinical setting/Golestan University of Medical Sciences/ 2006/
14. Sharif F, masoumi S/ A qualitatative study of nursing

    student experiences of clinical practice/ BMC Nursing 2005;4:6/
15. Mortaz Hejri S, Jalili M/ Analyzing the Quality of Objective Structured Clinical Examination in Tehran University of Medical Sciences/ Iran J Med Educ/ 2017;17:50–9/
16. Arnold L/ Assessing professional behavior: yesterday, today, and tomorrow/ Acad Med/ 2002;77(6):502–15/
17. Aranda S, Yates P/ A National Professional Development Framework for Cancer Nursing (2nd edn)/ Canberra: The National Cancer Nursing Education Project (EdCaN), Cancer Australia/ 2009
18. Tabrizi JS/ Objective Structured Field Examination (OSFE) TT/ IJME [Internet]/ 2012 Mar 1;11(8):976–8/ Available from: http://ijme/mui/ac/ir/article-1-1474-en/html
19. Al-Wardy NM/ Assessment methods in undergraduate medical education/ Sultan Qaboos Univ Med J [Internet]/ 2010/07/19/ 2010 Aug;10(2):203–9/ Available from: https://pubmed/ncbi/nlm/nih/gov/21509230
20. Pell G, Fuller R, Homer M, Roberts T/ How to measure the quality of the OSCE: a review of metrics–AMEE guide no/ 49/ Med Teach/ 2010;32(10):802–11/
21. Fuller R, Homer M, Pell G/ Longitudinal interrelationships of OSCE station level analyses, quality improvement and overall reliability/ Med Teach/ 2013;35(6):515–7/
22. . Homer M, Pell G, Fuller R, Patterson J. Quantifying error in OSCE standard setting for varying cohort sizes: A resampling approach to measuring assessment quality. Med Teach. 2016; 38(2):181-8.
23. ۲۳- MortazHejri S, Jalili M, Muijtjens AMM, Van Der Vleuten CPM. Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. J Res Med Sci. 2013; 18(10): 887–891.
24. Huang CC, Chan CY, Wu CL, Chen YL, Yang HW, Huang CC, Chen CH, Huang WJ, Lee FY, Lin SJ, Chen JW. Assessment of clinical competence of medical students using the objective structured clinical examination: first 2 years' experience in Taipei Veterans General Hospital. Journal of the Chinese Medical Association. 2010 Nov 1;73(11):589-95.
25. Elabd K, Abdul-Kadir H, Alkhenizan A, Alkhalifa MK. A Comparison of the Checklist Scoring Systems, Global Rating Systems, and Borderline Regression Method for an Objective Structured Clinical Examination for a Small Cohort in a Saudi Medical School. Cureus. 2023 Jun 5;15(6).
26. Park SY, Lee SH, Kim MJ, Ji KH, Ryu JH. Comparing the cut score for the borderline group method and borderline regression method with norm-referenced standard setting in an objective structured clinical examination in medical school in Korea. Journal of educational evaluation for health professions. 2021 Sep 27;




[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]

AUMJ Alborz University of Medical Sciences Journal

*Alborz University of Medical Sciences Journal, Fall 2023, Vol. 12, No. 4:485 - 495* *Original Article*

**Suleiman Ahmadi[1], Amin Habibi[2], Mitra rahimzadeh[3], Shahla Bahrami[4*]**

*[1]Department of medical education, Virtual School of Medical Education, Shahid Beheshti university of Medical sciences, Tehran, Iran.*
*[2]Department of medical education, smart university of Medical sciences, Tehran, Iran.*
*[3]Social Determinants of Health Research Center, Biostatistics and Epidemiology Department. Public Health School, Alborz,Iran.*
*[4]Masters student in medical education, Virtual School of Medical Education, Shahid Beheshti university of Medical sciences, Tehran, Iran.*

# Analyzing the Quality of Objective Structured Clinical Examination in Alborz University of Medical Sciences

Received: 10 Jan 2023 ; Accepted: 18 Sep 2023

## Abstract

**Introduction:** One of the evaluation methods is based on the performance of objective structured clinical examination. Passing this exam is necessary for medical students to graduate, practice medicine and participate in a specialized course. This study is designed to analyze the quality of these tests using psychometric indicators.

**Methods:** In this cross-sectional descriptive study, the test of three periods of clinical competence, respectively, were analyzed. Average (standard deviation), the cut-off score was (by borderline regression method), rejection rate, R2 determination coefficient, root mean square error to estimate the the cut-off score, difficulty coefficient, neatness coefficient were calculated in individual stations and the whole test.

**Findings:** 64 medical students along with 26 expert examiners from each station were present in this exam. According to the cut-off score in the 31st exam, all students got a passing grade and we did not fail. The total RMSE was equal to 0.55 and the value of $R_2$ in different stations was obtained from 0.46 to 0.85. The hardest and easiest stations based on the difficulty coefficient were gynecology, social medicine and orthopedics respectively. In the 32nd exam, the rejection rate was one person (0.23%). The total RMSE was 0.57 and the value of $R_2$ in different stations was obtained from 0.38 to 0.87. The most difficult and the easiest stations based on the difficulty coefficient were radilogy, emergency medicine and surgery respectively. In the 33rd exam, all students got a passing grade. Total RMSE was equal to 0.76 and $R_2$ value was obtained in different stations from 0.39 to 0.95. The most difficult and easiest stations based on the difficulty coefficient were internal glands and surgery respectively.

**Conclusion**: To improve the quality of the OSCE exam, measuring its quality is vital. The use of several psychometric indicators can be an important step in determining the quality of the test. The results obtained during the three test periods indicate that the quality of the test is acceptable in most of the stations and some cases need to be reviewed and improved.

**Keywords:** Test analysis, Objective Structured Clinical Examination, clinical competency, psychometrics, border Regression

***Corresponding Author:**
Masters student in medical education, Virtual School of Medical Education, Shahid Beheshti university of Medical sciences, Tehran, Iran.

09901868707
bahrami3636@gmail.com

نشریه علمی پژوهشی دانشگاه علوم پزشکی البرز، پاییز ۱۴۰۲؛ دوره ۱۲، شماره ۴ ؛ ۴۸۵ تا ۴۹۵

[ Downloaded from aums.abzums.ac.ir on 2026-10-30 ]

[ DOI: 10.61186/aums.12.4.485 ]