

# بررسی روند تغییرات کیفیت آزمون عینی ساختارمند بالینی برای سنجش صلاحیت بالینی دانشجویان پزشکی در دانشگاه علوم پزشکی تهران

زینب غفاری

دستیار داخلی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

دکتر محبوبه خباز مافی نژاد

دانشیار، گروه آموزش پزشکی، مرکز تحقیقات آموزش علوم پزشکی، مرکز مطالعات و توسعه آموزش، دانشگاه علوم پزشکی تهران، تهران، ایران

فاطمه شهبازی

مربی، گروه پزشکی اجتماعی، دانشکده پزشکی، مرکز مطالعات و توسعه آموزش، دانشگاه علوم پزشکی زاهدان، زاهدان، ایران

آرش عکاظی

گروه پزشکی قانونی، دانشگاه علوم پزشکی تهران، تهران، ایران

دکتر احسان باستان حق<sup>1</sup>

دانشیار، گروه بیهوشی، دانشگاه علوم پزشکی تهران، تهران، ایران

## Investigating the Trend of Changes in the Quality of Objective Structured Clinical Examination to Assess the Clinical Competence of Medical Students in Tehran University of Medical Sciences

**Zeinab Ghaffari**

Assistant Internal Medicine, Shahid Beheshti University of medical sciences, Tehran, Iran

**Mahboobeh Khabaz Mafinejad, MD**

Associate Professor, Department of Medical Education, Health Professions Education Research Center, Education Development Center (EDC), Tehran University of Medical Sciences, Tehran, Iran

**Fatemeh Shahbazi**

Instructor of Medical Education, Department of Community Medicine, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran.

**Arash Okazi**

Department of Forensic Medicine, Tehran University of Medical Sciences, Tehran, Iran

**Ehsan Bastanhagh, MD**

Associate Professor, Department of Anesthesiology, Tehran University of Medical Sciences, Tehran, Iran

### ABSTRACT

**Introduction:** Considering that in recent years the clinical competency test is performed to measure the performance of medical students before graduation, it is necessary to study its psychometric properties in order to make managerial decisions to hold a valid and reliable test.

**Materials and Methods:** In this retrospective study, all medical students of Tehran University of Medical Sciences who participated in the clinical competency test during the academic year 2017-2018 were included in the study. Participants' scores and the total score of the OSCE test were collected and analyzed separately for each station. In order to evaluate the quality of OSCE, Difficulty Index, Discrimination Index,  $R^2$  and differentiation of degrees were calculated using SPSS software. The reliability of the test was determined using Cronbach's alpha. The correlation between the total grade point average of the students was calculated by the final OSCE score.

**Results:** 772 interns were included in the study in the form of 4 courses of competency test. The value of  $R^2$  for each test is 0.33 to 0.78%, 0.75 to 0.91%, 0.63 to 0.90%, 0.53 to 0.89%, respectively. The Cronbach's alpha test was obtained from 0.52 to 0.76. The range of Discrimination Index for each test was equal to -0.02 to 0.52, -0.09 to 0.64,

<sup>1</sup>. نویسنده مسؤول: e-bastanhagh@sina.tums.ac.ir

order to evaluate the quality of OSCE, Difficulty Index, Discrimination Index,  $R^2$  and differentiation of degrees were calculated using SPSS software. The reliability of the test was determined using Cronbach's alpha. The correlation between the total grade point average of the students was calculated by the final OSCE score.

**Results:** 772 interns were included in the study in the form of 4 courses of competency test. The value of  $R^2$  for each test is 0.33 to 0.78%, 0.75 to 0.91%, 0.63 to 0.90%, 0.53 to 0.89%, respectively. The Cronbach's alpha test was obtained from 0.52 to 0.76. The range of Discrimination Index for each test was equal to -0.02 to 0.52, -0.09 to 0.64, -0.05 to 0.64, and 0.94 to 1 respectively. The range of Difficulty Index for each test was 0.68 to 0.93 and 0.21 to 0.88, 0.62 to 0.91, and 0.72 to 0.96, respectively. The correlation coefficient between the total grade point average and the final OSCE score was calculated to be 0.56 ( $P < 0.001$ ).

**Conclusion:** The internal consistency of the tests was acceptable. The R-index varied at different test times at different stations. The range of Difficulty Index and Discrimination Index in different test times have changes that can be resulting from the educational effects of test implementation for students and improving the design quality of stations. There was a direct and positive relationship between students' GPA and OSCE final score based on Pearson correlation coefficient.

**Keywords:** competence, OSCE, medicine, psychometric

## چکیده

**مقدمه:** با توجه به آنکه در سال‌های اخیر آزمون صلاحیت بالینی به منظور سنجش عملکرد دانشجویان پزشکی پیش از فارغ‌التحصیلی اجرا می‌شود، بررسی ویژگی‌های سایکومتریک آن به منظور اتخاذ تصمیمات مدیریتی در زمینه برگزاری یک آزمون روا و پایا ضروری است.

**مواد و روش‌ها:** در این مطالعه گذشته‌نگر، تمامی دانشجویان پزشکی دانشگاه علوم پزشکی تهران که طی سال تحصیلی ۲۰۱۷-۲۰۱۸ در آزمون صلاحیت بالینی شرکت کردند، وارد مطالعه شدند. نمرات شرکت‌کنندگان به تفکیک ایستگاه‌ها و نمره کل آزمون آسکی جمع‌آوری و تحلیل شد. به منظور بررسی کیفیت آزمون آسکی صلاحیت بالینی، ضریب دشواری، ضریب تمیز،  $R^2$  و تمایز بین درجات با استفاده از نرم‌افزار SPSS محاسبه شد. پایایی آزمون با استفاده از آلفای کرونباخ تعیین شد. همبستگی بین معدل کل دانشجویان با نمره نهایی آسکی محاسبه شد.

**یافته‌ها:** تعداد ۳۲۷ نفر کارورز در قالب ۴ دوره برگزاری آزمون صلاحیت وارد مطالعه شدند. مقدار  $R^2$  برای هر آزمون به ترتیب از ۰.۳۳ تا ۰.۷۸، ۰.۷۵ تا ۰.۹۱، ۰.۶۳ تا ۰.۹۰ و ۰.۵۳ تا ۰.۸۹ به دست آمد. مقدار آلفای کرونباخ آزمون از ۰.۵۲ تا ۰.۷۶ به دست آمد. دامنه ضریب تمیز به ترتیب برای هر آزمون برابر ۰/۰۰۲- تا ۰.۵۲، ۰.۰۹- تا ۰.۶۴، ۰.۰۵- و ۰.۹۴ تا ۱ بود. دامنه ضریب دشواری به ترتیب برای هر آزمون برابر ۰.۶۸ تا ۰.۹۳، ۰.۲۱ تا ۰.۸۸، ۰.۶۲ تا ۰.۹۱ و ۰.۷۲ تا ۰.۹۶ بود. ضریب همبستگی بین معدل کل و نمره نهایی آسکی صلاحیت بالینی ۰.۵۶ ( $P < 0.001$ ) محاسبه شد.

**نتیجه‌گیری:** ضرایب تحلیل پس‌آزمون آسکی در نوبت‌های مختلف اجرا می‌تواند به دلیل اثرات آموزشی اجرای آزمون بر یادگیری دانشجویان و بهبود کیفیت طراحی ایستگاه‌ها تغییر کند. محاسبه شاخص‌های سایکومتریک متناسب با آزمون‌های مبتنی بر عملکرد برای تحلیل ایستگاه‌ها و نیز تحلیل کل آزمون قابل اجرا است و اطلاعات مفیدی در مورد نقاط مثبت و منفی آزمون در اختیار می‌گذارد که برای ارائه راهکارهای اصلاحی حائز اهمیت است.

**کلواژگان:** صلاحیت، آزمون آسکی، پزشکی، کیفیت آزمون

## مقدمه

و مخدوش کننده در روند معاینه و تشخیص، اشاره کرد (۱۳) و اگرچه آزمون عینی ساختارمند بالینی از ابزارهای ارزیابی متداول است و اثرات آموزشی بالایی به همراه دارد، اما اجرای صحیح آن به دلیل نیاز به منابع انسانی کمی دشوار است و با مشکلاتی از قبیل سخت‌گیری یا سهل‌گیری آزمون‌گران، تفاوت در عملکرد بیمارناها، تفاوت در سطح دشواری ایستگاه‌ها، اضطراب بالای دانشجویان و برگزار کنندگان، مشکلات مربوط به نمره‌دهی به دلیل تعدد آزمون-گران، همراه است (۲۰-۲۸). با توجه به آن که این آزمون در سال‌های اخیر به عنوان یک آزمون سرنوشت‌ساز در انتهای دوره آموزشی و با هدف سنجش صلاحیت بالینی دانشجویان پزشکی پیش از فارغ‌التحصیل شدن اجرا می‌شود و همچنین در آزمون‌هایی با این ساختار، عوامل متعددی وجود دارند که می‌توانند سبب ایجاد خطا شوند، بنابراین سنجش مداوم کیفیت نحوه برگزاری این آزمون امری حیاتی به شمار می‌رود و بسیاری از مراکز آموزشی از جمله دانشگاه علوم پزشکی تهران در پی رفع نواقص و ارتقا آن هستند (۲۹ و ۳۰). بررسی روند تغییرات ویژگی‌های سایکومتریک آزمون عینی ساختارمند بالینی با هدف سنجش صلاحیت بالینی دانشجویان پزشکی به منظور اتخاذ تصمیمات مدیریتی در زمینه برگزاری یک آزمون روا و پایا بسیار حائز اهمیت است. به همین منظور هدف اصلی از این مطالعه، بررسی روند تغییرات کیفیت آزمون عینی ساختارمند بالینی (آسکی<sup>۲</sup>) برای سنجش صلاحیت بالینی دانشجویان پزشکی در دانشگاه علوم پزشکی تهران است.

## مواد و روش‌ها

در این مطالعه گذشته‌نگر، ۳۲۷ نفر از دانشجویان پزشکی دانشکده پزشکی دانشگاه علوم پزشکی تهران که در سال ۱۳۹۷ در امتحان صلاحیت بالینی شرکت کرده بودند، وارد مطالعه شدند. آزمون آسکی صلاحیت بالینی در سال ۱۳۹۷ در ۴ نوبت متوالی (خرداد ماه، مرداد ماه، آبان ماه و بهمن ماه) برای دانشجویان مقطع کارورزی که در شرف فارغ‌التحصیلی بودند، در دانشگاه علوم پزشکی تهران برگزار شد. در هر یک از آزمون‌های آسکی صلاحیت بالینی ۱۴ ایستگاه برای سنجش صلاحیت دانشجویان پزشکی شرکت کننده وجود

ارزیابی عملکرد دانشجویان پزشکی همواره به عنوان یکی از حساس‌ترین وظایف پیش روی مؤسسات آموزشی به شمار می‌آید (۱). به این منظور متخصصان آموزش پزشکی در جستجوی روش‌های معتبری بوده‌اند که به طور مؤثر دانش و کارایی بالینی دانشجویان را ارزیابی کنند (۲). مرور شواهد نشان می‌دهد با وجود روش‌های متعدد ارزشیابی بالینی، به طور معمول ارزیابی دانشجویان محدود به اطلاعات ذهنی بوده نه سنجش مهارت‌های بالینی و عملی دانشجویان (۳). از بین آزمون‌های مختلف معرفی شده، آزمون عینی ساختارمند بالینی با ارائه نتایج رضایت‌بخش، فرآیند ارزیابی آموزشی را متحول کرده است (۴). به علاوه در مقایسه با آزمون موارد بالینی سنتی، از عینیت و پایایی بهتری برای ارزیابی مهارت‌های بالینی برخوردار است (۵). آزمون عینی ساختارمند بالینی در طول سی سال گذشته به یکی از ابزارهای ارزیابی متداول در آموزش پزشکی تبدیل شده است (۶ و ۷). اهمیت این آزمون از این جهت است که در صورت اجرای اصولی آزمون به ویژه برای دانشجویان سال آخر پزشکی، افرادی که از حداقل صلاحیت‌های مورد انتظار برخوردار هستند شناسایی می‌شوند (۸-۱۱). در این آزمون، موقعیت‌هایی مشابه با شرایط واقعی بالینی در قالب ایستگاه‌های متعدد طراحی می‌شود (۱۲ و ۱۳) که طی آن توانایی و مهارت خاصی از دانشجو مورد ارزیابی قرار می‌گیرد (۱۴). در واقع در این آزمون به جای بررسی صرف یک مهارت خاص، چندین مهارت مورد ارزیابی قرار می‌گیرد (۱۳). از جمله مهارت‌های مورد سنجش در این آزمون می‌توان به سنجش مهارت شرح حال گرفتن، معاینه بالینی، تشخیص، انجام پروسیجرهای تشخیصی و درمانی، مهارت ارتباطی و آموزش به بیمار و ... اشاره کرد (۱۵ و ۱۶). از جمله مزایای آزمون عینی ساختارمند بالینی می‌توان به یکسان بودن شرایط ارزیابی برای همه آزمون‌شوندگان، فرصت ارزیابی مهارت‌های متعدد داوطلب در یک آزمون، سنجش عینی عملکرد دانشجو با استفاده از چک لیست‌های مشخص، و جایگزینی بیمارنا با بیماران واقعی با هدف کاهش بروز خطاهای احتمالی ناشی از عدم همکاری مناسب بیماران واقعی و یا وجود علائم غیر طبیعی

درجات<sup>۱۱</sup> از طریق معادله رگرسیون برای هر ایستگاه محاسبه شد. در محاسبه  $R^2$  نمرات چک‌لیست از بین اعداد صفر تا ده و نمره گلوبال از بین کلمات مردود، مرزی، قابل قبول، خوب و عالی انتخاب شد و ارتباط بین تغییرات نمره چک‌لیست هر ایستگاه (متغیر وابسته) با نمره گلوبال کیفی (متغیر مستقل) بررسی شد. اگر  $R^2 > 0.5$  باشد یعنی رابطه منطقی بین نمرات چک‌لیست و گلوبال وجود دارد. به عبارت دیگر، ۵۰ درصد تغییرات در نمرات گلوبال به دلیل تغییرات در نمره چک‌لیست آنها بوده و با تفاوت‌های موجود در نمرات چک‌لیست قابل توجیه است. برخی از ایستگاه‌ها صرفاً به شکل پاسخگویی به سؤالات کتبی و بدون حضور فرد آزمونگر برگزار شده بود که در نتیجه ایستگاه‌های مذکور فاقد نمره گلوبال بودند و ضریب  $R^2$  مابقی ایستگاه‌ها از طریق معادله رگرسیون محاسبه گردید. برای محاسبه ضریب دشواری، نمره خام دانشجویان (متغیر پیوسته) به عنوان نتیجه ایستگاه در نظر گرفته شد و همانند سؤالات تشریحی، نمره میانگین ایستگاه بر نمره کل آن (۱۰) تقسیم شد تا ضریب دشواری محاسبه گردد.

### یافته‌ها

از مجموع شرکت کنندگان، ۲۰۱ نفر (۶۱.۵ درصد) زن و ۱۲۶ نفر (۳۸.۷ درصد) مرد بودند. میانگین سنی افراد ۲۶.۱۹ سال با انحراف معیار ۱.۹۳ بود. دامنه سنی شرکت کنندگان بین ۲۳ تا ۴۶ سال قرار داشت. میانگین معدل کل دانشجویان ۱۶.۳۹ و میانگین نمره نهایی آسکی ۱۶.۰۵ بود. در این مطالعه، داده‌های مربوط به معدل کل و نمره نهایی آزمون صلاحیت بالینی که به ترتیب در خرداد، مرداد، آبان و بهمن ماه (۴ نوبت) در سال ۱۳۹۶ برگزار شد در جدول شماره یک ارائه شده است. ضریب همبستگی پیرسون برای دو متغیر پیوسته معدل کل و نمره نهایی آسکی دانشجویان مقدار ۰.۵۶ با ( $P < 0.001$ ) محاسبه شد و به این معنا است که بین این دو متغیر رابطه مستقیم مثبت متوسط وجود دارد. برای آزمون‌های مورد بررسی به ترتیب آلفای کرونباخ آزمون خرداد ماه ۰.۵۲، مرداد ماه ۰.۷۶، آبان ماه ۰.۷۵ و بهمن ماه ۰.۶۰ محاسبه گردید. دامنه تغییرات آلفا در صورت حذف ایستگاه به ترتیب در هر نوبت اجرای آزمون بین ۰.۳۸ تا ۰.۵۶، ۰.۷۱ تا ۰.۷۶، ۰.۷۱ تا ۰.۷۵ و ۰.۵۳ تا ۰.۶۱ متغیر بود. خطای معیار اندازه‌گیری (SEM) آزمون آسکی صلاحیت بالینی به

داشت. مدت زمان هر یک از ایستگاه‌های آزمون صلاحیت بالینی، ۵ دقیقه بود که در مجموع مدت زمان کلی هر آزمون آسکی ۷۰ دقیقه در نظر گرفته شد. به منظور حفظ محرمانیت اطلاعات، نمرات هر یک از دانشجویان بدون ذکر نام و با ارائه کد بررسی شد. برای انجام این مطالعه از برگه‌های جمع‌آوری داده‌ها و همچنین پرسشنامه جهت جمع‌آوری و ثبت داده‌ها استفاده شد. اطلاعات از معاونت آموزشی دانشکده و مرکز مهارت‌های بالینی دانشگاه جمع‌آوری شد. به منظور تعیین خصوصیات سایکومتریک آزمون آسکی، ضریب پایایی کل آزمون محاسبه شد و برای تعیین ویژگی‌های سایکومتریک ایستگاه‌ها بر پایه نظریه کلاسیک آزمون، دو موضوع مدنظر قرار گرفت. اول اینکه آزمون آسکی به عنوان یک آزمون مبتنی بر معیار در نظر گرفته شد. به این معنا که بیشتر از اینکه نمره خام دانشجویان در مقایسه با یکدیگر اهمیت داشته باشد، رسیدن/ نرسیدن به حد تسلط و در نتیجه قبول و رد داوطلبان در آزمون اهمیت دارد. در نتیجه در تحلیل پس‌آزمون از فرمول‌های معمول که برای آزمون‌های مبتنی بر هنجار استفاده می‌شود، اجتناب و از فرمول‌های مبتنی بر معیار استفاده شد. ثانیاً نتیجه هر ایستگاه به دو صورت نمره خام (متغیر پیوسته) و نتیجه رد و قبول (متغیر گسسته دو ارزشی) بیان شد. در این مطالعه هر دو حالت در نظر گرفته شد و در نهایت مقایسه بین دو روش انجام شد.

تجزیه و تحلیل نمرات شرکت کنندگان به تفکیک ایستگاه‌ها و نمره کل آزمون آسکی محاسبه شد. برای محاسبه با توجه به آنکه نمرات چک‌لیست‌های مختلف با یکدیگر برابر نبودند، نمرات تمام ایستگاه‌ها از نمره ۱۰ و نمره نهایی آزمون از ۲۰ محاسبه شد. نمره حد نصاب به صورت مقدار ثابت ۵۰ درصد کل نمره آزمون (نمره ۱۰) در نظر گرفته شده است. به منظور بررسی کیفیت آزمون آسکی صلاحیت بالینی، ضریب دشواری، ضریب تمیز،  $R^2$ ، SEM، تمایز بین درجات و پایایی با استفاده از نرم‌افزار SPSS محاسبه شد. پایایی آزمون با استفاده از آلفای کرونباخ محاسبه شد. سپس با در نظر گرفتن پایایی، خطای معیار اندازه‌گیری (SEM) آزمون‌ها به دست آمد. در مطالعه حاضر شاخص آلفای کرونباخ در صورت حذف هر ایستگاه به جهت بررسی کیفیت محتوای ایستگاه محاسبه گردید. علاوه بر این، دو شاخص "ضریب  $R^2$ " و "تمایز بین

مدت زمان اختصاص داده شده به هر یک از ایستگاه‌ها و مدت زمان کل آزمون بر کسب پایایی بالاتر آزمون تأثیرگذار است (۳۱). به نظر می‌رسد در آزمون‌های صلاحیت بالینی دانشگاه ما، با افزایش تعداد ایستگاه‌ها و همچنین طراحی ایستگاه‌ها بر طبق بلوپرینت آزمون، شاخص آلفای کرونباخ در نتیجه افزایش تعداد ایستگاه‌ها افزایش می‌یابد. در مطالعه مرتاض هجری و همکاران این موضوع تأیید شد (۱۵). به صورت کلی مقدار شاخص  $R^2$  متوسط یعنی بالای ۰.۵ نشان دهنده رابطه منطقی بین نمرات گلوبال و چکلیست است (۲). در مطالعه ما، این شاخص در ایستگاه‌های آزمون خرداد ماه در سه ایستگاه اطفال (توصیه‌های بعد از واکسیناسیون)، درد قفسه سینه و تعبیه لوله بینی معدی به ترتیب ۰.۴۴، ۰.۴۲ و ۰.۳۳ بود که از حد مطلوب پایین‌تر بود. به عبارت دیگر برخی از دانشجویان علی‌رغم اینکه نمره چکلیست خوبی داشتند، وضعیت آنها در ارزیابی گلوبال رضایت‌بخش نبود و برعکس. این مسأله می‌تواند نشان دهنده عدم درک صحیح آزمونگر از ارزیابی کلی عملکرد یا مشکل در چکلیست طراحی شده بوده و از تفاوت برداشت آزمونگر و طراح چکلیست ناشی شده باشد (۲). به این منظور به نظر می‌رسد در آزمون‌های مشابه دیگر می‌توان از طریق ارائه برنامه‌های توانمندسازی برای ارزیابان و همچنین بازنگری چکلیست‌های ارزیابی به بهبود شاخص  $R^2$  در ایستگاه‌ها کمک کرد.

ترتیب ۰.۸۱، ۰.۸۰، ۰.۸۲ و ۰.۷۴ محاسبه شد. جزئیات اطلاعات مربوط به محاسبه ضریب  $R^2$  و شاخص تمایز بین درجات برای ایستگاه‌های آزمون در ۴ دوره متوالی در جدول شماره ۲ ارائه شده است.

دامنه ضریب دشواری به ترتیب در ۴ دوره متوالی آزمون برابر ۰.۶۸ تا ۰.۹۳، ۰.۲۱ تا ۰.۸۸، ۰.۶۲ تا ۰.۹۱ و ۰.۷۲ تا ۰.۹۶ بود. اطلاعات مربوط به ضریب دشواری و ضریب تمیز هر یک از ایستگاه‌های آزمون آسکی در جدول شماره ۳ ارائه شده است.

### بحث

روش‌های متعددی برای ارزیابی صلاحیت بالینی دانشجویان وجود دارد. از جمله روش‌های ارزیابی متداول، آزمون عینی ساختارمند بالینی (آسکی) است (۴). در این مطالعه شاخص‌های سایکومتریک آزمون آسکی صلاحیت بالینی دانشجویان پزشکی در دانشگاه علوم پزشکی تهران در سال ۱۳۹۶ در سطح کل آزمون و در سطح هر یک از ایستگاه‌ها مورد بررسی قرار گرفت.

به منظور بررسی ثبات درونی آزمون و کیفیت هر ایستگاه، آلفای کرونباخ محاسبه شد که در نوبت اول اجرای آزمون، از پایایی پایین‌تری در مقایسه با نوبت‌های بعدی برخوردار بود. در مطالعه مروری که توسط برانیک و همکاران در مورد پایایی آسکی انجام شد، ۳۹ مقاله مورد بررسی قرار گرفت و میانگین آلفای کرونباخ به صورت کلی ۰.۶۶ به دست آمد (۳۱). وجود عواملی از قبیل افزایش تعداد ایستگاه‌های آزمون، افزایش

جدول ۱: اطلاعات مربوط به معدل کل و نمره نهایی آزمون صلاحیت بالینی در ۴ نوبت متوالی

| نوبت | آزمون                 | تعداد شرکت‌کنندگان | حداکثر نمره | حداقل نمره | میانگین | انحراف معیار |
|------|-----------------------|--------------------|-------------|------------|---------|--------------|
| ۱    | نمره نهایی آزمون آسکی | ۶۷                 | ۱۷.۲۲       | ۱۲.۷۲      | ۱۵.۶۶   | ۰.۸۷         |
| ۲    | معدل کل               |                    | ۱۸.۶۸       | ۱۳.۳۸      | ۱۶.۱۷   | ۱.۱۹         |
| ۳    | نمره نهایی آزمون آسکی | ۸۰                 | ۱۸.۶۲       | ۱۲.۰۴      | ۱۵.۴۸   | ۱.۶۷         |
| ۴    | معدل کل               |                    | ۱۸.۹۷       | ۱۳.۵۳      | ۱۶.۳۶   | ۱.۳۸         |
| ۵    | نمره نهایی آزمون آسکی | ۷۶                 | ۱۸.۷۹       | ۱۰.۷۹      | ۱۵.۹۱   | ۱.۱۸         |
| ۶    | معدل کل               |                    | ۱۸.۵۴       | ۱۳.۹۱      | ۱۶.۵۷   | ۱.۱۷         |
| ۷    | نمره نهایی آزمون آسکی | ۱۰۴                | ۱۹.۲۵       | ۱۲.۳۳      | ۱۷.۱۶   | ۱.۰۹         |
| ۸    | معدل کل               |                    | ۱۸.۹۶       | ۱۴.۰۱      | ۱۶.۴۸   | ۰.۹۶         |

جدول ۲: ضریب آلفای در صورت حذف ایستگاه، ضریب  $R^2$  و تمایز بین درجات ایستگاه‌های آزمون در ۴ نوبت متوالی

| ایستگاه       | نمره آلفا در صورت حذف ایستگاه | ضریب $R^2$ | تمایز بین درجات | ایستگاه | نمره آلفا در صورت حذف ایستگاه | ضریب $R^2$ | تمایز بین درجات | ایستگاه               | نمره آلفا در صورت حذف ایستگاه | ضریب $R^2$ | تمایز بین درجات |
|---------------|-------------------------------|------------|-----------------|---------|-------------------------------|------------|-----------------|-----------------------|-------------------------------|------------|-----------------|
|               |                               |            |                 |         |                               |            |                 |                       |                               |            |                 |
| خرداد         | روانپزشکی                     | ۰/۷۶       | ۰/۸۴            | ۱/۸۲    | روانپزشکی                     | ۰/۷۳       | ۱/۱۹            | روانپزشکی             | ۰/۵۲                          | ۰/۷۳       | ۱/۱۹            |
|               | نورولوژی                      | ۰/۷۳       | ۰/۸۵            | ۱/۵۳    | نورولوژی                      | ۰/۷۵       | ۱/۱۷            | نورولوژی              | ۰/۴۹                          | ۰/۷۵       | ۱/۱۷            |
|               | ارتوپدی                       | ۰/۷۴       | ۰/۸۰            | ۱/۴۱    | ارتوپدی                       | ۰/۷۸       | ۰/۹۸            | ارتوپدی               | ۰/۴۸                          | ۰/۷۸       | ۰/۹۸            |
|               | جراحی                         | ۰/۷۳       | ۰/۸۹            | ۱/۴۳    | اورولوژی                      | ۰/۶۱       | ۰/۸۲            | اورولوژی              | ۰/۴۸                          | ۰/۶۱       | ۰/۸۲            |
|               | بیهوشی                        | ۰/۷۲       | ۰/۸۸            | ۱/۶۲    | اورژانس                       | ۰/۶۶       | ۱/۰۹            | اورژانس               | ۰/۳۸                          | ۰/۶۶       | ۱/۰۹            |
|               | اورژانس                       | ۰/۷۵       | ۰/۹۱            | ۲/۰۲    | پوست                          | ۰/۵۷       | ۱/۰۲            | پوست                  | ۰/۵۶                          | ۰/۵۷       | ۱/۰۲            |
|               | داخلی                         | ۰/۷۴       | ۰/۸۳            | ۱/۹۰    | اطفال                         | ۰/۴۴       | ۰/۷۹            | اطفال                 | ۰/۵۳                          | ۰/۴۴       | ۰/۷۹            |
|               | اطفال                         | ۰/۷۴       | ۰/۸۲            | ۱/۹۶    | زنان                          | ۰/۵۱       | ۰/۹۸            | زنان                  | ۰/۴۶                          | ۰/۵۱       | ۰/۹۸            |
|               | زنان                          | ۰/۷۱       | ۰/۸۶            | ۱/۷۷    | لوله گذاری بینی-معدده         | ۰/۳۳       | ۰/۶۳            | لوله گذاری بینی-معدده | ۰/۵۲                          | ۰/۳۳       | ۰/۶۳            |
|               | قلب                           | ۰/۷۳       | ۰/۸۷            | ۱/۵۴    | فشارخون                       | ۰/۶۹       | ۱/۱             | فشارخون               | ۰/۵۰                          | ۰/۶۹       | ۱/۱             |
|               | عفونی                         | ۰/۷۴       | ۰/۷۵            | ۱/۵۹    | درد قفسه سینه                 | ۰/۴۲       | ۱/۲             | درد قفسه سینه         | ۰/۴۵                          | ۰/۴۲       | ۱/۲             |
|               | -                             | -          | -               | -       | اسپری                         | ۰/۶۹       | ۱/۳۹            | اسپری                 | ۰/۵۱                          | ۰/۶۹       | ۱/۳۹            |
|               | اردیبهشت                      | روانپزشکی  | ۰/۵۷            | ۰/۶۷    | ۱/۹۴                          | روانپزشکی  | ۰/۶۳            | ۲/۲۲                  | روانپزشکی                     | ۰/۷۴       | ۰/۶۳            |
| ارتوپدی       |                               | ۰/۵۴       | ۰/۷۲            | ۱/۲۰    | نورولوژی                      | ۰/۸۱       | ۱/۸۱            | نورولوژی              | ۰/۷۲                          | ۰/۸۱       | ۱/۸۱            |
| اورولوژی      |                               | ۰/۶۰       | ۰/۶۱            | ۰/۸۸    | ارتوپدی                       | ۰/۸۴       | ۱/۴۱            | ارتوپدی               | ۰/۷۲                          | ۰/۸۴       | ۱/۴۱            |
| جراحی         |                               | ۰/۵۹       | ۰/۵۳            | ۰/۷۱    | اورولوژی                      | ۰/۸۴       | ۱/۲۵            | اورولوژی              | ۰/۷۲                          | ۰/۸۴       | ۱/۲۵            |
| اورژانس       |                               | ۰/۵۸       | ۰/۸۸            | ۱/۷۹    | جراحی                         | ۰/۸۱       | ۱/۳۷            | جراحی                 | ۰/۷۳                          | ۰/۸۱       | ۱/۳۷            |
| گوش، حلق بینی |                               | ۰/۶۰       | ۰/۸۴            | ۱/۴۸    | بیهوشی                        | ۰/۸۹       | ۱/۶۹            | بیهوشی                | ۰/۷۳                          | ۰/۸۹       | ۱/۶۹            |
| بیهوشی        |                               | ۰/۵۸       | ۰/۸۹            | ۱/۶۷    | اورژانس                       | ۰/۸۰       | ۱/۵۷            | اورژانس               | ۰/۷۲                          | ۰/۸۰       | ۱/۵۷            |
| داخلی         |                               | ۰/۵۳       | ۰/۷۹            | ۱/۸۷    | داخلی                         | ۰/۸۱       | ۱/۷۱            | داخلی                 | ۰/۷۳                          | ۰/۸۱       | ۱/۷۱            |
| اطفال         |                               | ۰/۵۹       | ۰/۸۰            | ۱/۵۰    | اطفال                         | ۰/۸۷       | ۱/۸۷            | اطفال                 | ۰/۷۱                          | ۰/۸۷       | ۱/۸۷            |
| قلب           |                               | ۰/۶۱       | ۰/۶۶            | ۱/۲۳    | عفونی                         | ۰/۸۳       | ۱/۴۷            | عفونی                 | ۰/۷۳                          | ۰/۸۳       | ۱/۴۷            |
| چشم           |                               | ۰/۵۶       | ۰/۸۲            | ۱/۴۸    | -                             | -          | -               | -                     | -                             | -          | -               |

سؤالات به درستی پاسخ دهند، بنابراین دور از انتظار نیست که ضریب تمیز سؤالات چندان بالا نباشند (۲). در سه آزمون خرداد و مرداد و آبان ماه ایستگاه‌هایی با مقدار صفر و منفی وجود داشتند، به ویژه در مورد ایستگاه پوست که در هر سه دوره منفی بود. در تحلیل ضریب تمیز منفی می‌توان به مشکل در چک‌لیست ایستگاه‌های مربوطه، مبهم بودن سناریوهای ایستگاه‌ها، و تجربه ناکافی ارزیابان برای ارزیابی عملکرد دانشجویان در آن ایستگاه اشاره کرد.

فولر و همکاران در مطالعه خود به این نتیجه رسیدند که شاخص  $R^2$  سه ایستگاه که در سال ۲۰۰۷ دارای مقادیری پایین‌تر از ۰.۵ بود، با اعمال تغییراتی در چک‌لیست در استفاده مجدد سال بعد تغییرات مثبت داشتند (۳۲). در سایر ایستگاه‌ها در هر ۴ نوبت برگزاری آزمون‌ها این شاخص مطلوب بود.

از نظر ضریب تمیز، در آزمون‌های معیار محور از جمله آزمون آسکی مورد انتظار است که اکثر فراگیران به اکثر

| گروه اول        |             |           | گروه دوم              |             |           |
|-----------------|-------------|-----------|-----------------------|-------------|-----------|
| ایستگاه         | ضریب دشواری | ضریب تمیز | ایستگاه               | ضریب دشواری | ضریب تمیز |
| روانپزشکی       | ۰/۷۵        | -۰/۰۷     | روانپزشکی             | ۰/۶۸        | ۰/۵       |
| نورولوژی        | ۰/۷۱        | ۰/۲۳      | نورولوژی              | ۰/۸۱        | ۰         |
| ارتوپدی         | ۰/۸۸        | ۰         | ارتوپدی               | ۰/۸۵        | ۰         |
| جراحی           | ۰/۸۷        | ۰         | اورولوژی              | ۰/۸۲        | ۰/۵       |
| بیهوشی          | ۰/۸۰        | ۰         | اورژانس               | ۰/۹۲        | ۰         |
| اورژانس         | ۰/۶۷        | ۰/۴۸      | پوست                  | ۰/۷۲        | -۰/۰۲     |
| داخلی           | ۰/۷۷        | ۰/۲۳      | اطفال                 | ۰/۸۸        | ۰         |
| اطفال           | ۰/۷۳        | ۰/۳۲      | زنان                  | ۰/۷۱        | ۰         |
| زنان            | ۰/۶۷        | ۰/۶۴      | لوله گذاری بینی-معدده | ۰/۹۳        | ۰         |
| قلب             | ۰/۷۶        | ۰/۳۴      | فشارخون               | ۰/۸۳        | ۰         |
| عفونی           | ۰/۷۳        | ۰/۱۷      | درد قفسه سینه         | ۰/۸۳        | ۰         |
| رادیولوژی       | ۰/۲۱        | ۰/۵۲      | اسپری                 | ۰/۷۶        | ۰/۰۴      |
| گروه سوم        |             |           | گروه چهارم            |             |           |
| ایستگاه         | ضریب دشواری | ضریب تمیز | ایستگاه               | ضریب دشواری | ضریب تمیز |
| روانپزشکی       | ۰/۷۲        | ۰/۹۸      | روانپزشکی             | ۰/۶۶        | ۰/۵۱      |
| ارتوپدی         | ۰/۹۱        | ۱         | نورولوژی              | ۰/۶۳        | ۰/۵۹      |
| اورولوژی        | ۰/۹۴        | ۱         | ارتوپدی               | ۰/۹۱        | ۰/۱۲      |
| جراحی           | ۰/۹۶        | ۱         | اورولوژی              | ۰/۸۹        | ۰         |
| اورژانس         | ۰/۸۳        | ۰/۹۴      | جراحی                 | ۰/۹۱        | ۰         |
| گوش، حلق و بینی | ۰/۹۱        | ۱         | بیهوشی                | ۰/۷۷        | ۰/۳۳      |
| بیهوشی          | ۰/۸۱        | ۰/۹۶      | اورژانس               | ۰/۸۵        | ۰/۱۲      |

جدول شماره ۳: ضریب دشواری و ضریب تمیز هر ایستگاه آزمون در ۴ نوبت متوالی

رادیولوژی ارائه نمی‌شود. این موضوع می‌تواند بر عملکرد دانشجویان در ایستگاه رادیولوژی تأثیرگذار باشد. جدای از آن، بر طبق مطالعات اولین بار برگزاری یک آزمون، می‌تواند بر تقویت یادگیری مهارت‌های دانشجویان از طریق اثر آموزشی آزمون در آزمون‌های بعدی آسکی تأثیرگذار باشد (۳۳). به علاوه در آزمون آسکی صلاحیت بالینی انتظار می‌رود که دانشجویان با شرکت فعال در کلاس‌های آموزشی و انجام پروسیجرهای تشخیصی و درمانی طی دوره آموزشی بالینی بتوانند به کسب مهارت‌های بالینی از قبیل نحوه اخذ شرح حال، معاینه فیزیکی و ... برسند. از طرف دیگر با توجه به آنکه طراحی محتوای ایستگاه‌ها نیز منطبق بر بلوپرینت آزمون انجام می‌شود، بنابراین انتظار می‌رود در آزمون آسکی حاضر که به صورت معیار محور برگزار شده است، اکثر دانشجویان بتوانند منطبق با آموزش‌های ارائه شده در بخش‌های مختلف بالینی، بتوانند در اکثر ایستگاه‌های آزمون آسکی عملکرد قابل

نتایج حاصل از این مطالعه نشان داد که در آزمون دوره بهمن ماه با استفاده از راهکارهایی از قبیل برگزاری جلسات مرور همتایان توسط استادانی که مسؤول طراحی محتوای ایستگاه-ها بودند و همچنین برنامه‌ریزی برای طراحی اثربخش محتوای ایستگاه‌هایی که توانایی افتراق بین افراد ضعیف و قوی را داشته باشد، تمامی ایستگاه‌ها دارای ضریب تمیز بالای ۰.۹ بودند.

بر اساس نتایج حاصل از مطالعه ما، در آزمون آسکی خرداد ماه که اولین دوره برگزاری آزمون صلاحیت بالینی بود، ایستگاه رادیولوژی پایین‌ترین ضریب دشواری را داشت. یک تفسیر برای این موضوع آن است که آموزش روتیشن رادیولوژی، در مقطع کارآموزی به صورت دوره یک ماهه فشرده در دانشگاه علوم پزشکی تهران برگزار می‌شود و در طول دوره کارروزی در مورد بخش‌های مختلف بالینی که نزدیک به زمان برگزاری آزمون صلاحیت بالینی است، آموزش

آسکی صلاحیت بالینی، به جای بررسی تنها یک دوره از آزمون بود. یکی از محدودیت‌های مطالعه استفاده از عدد ثابت ۱۰ به عنوان حد نصاب قبولی ایستگاه‌ها بود، در صورتی که اگر حد نصاب با روش‌های دیگری از جمله رگرسیون مرکزی، آنگوف تعدیل یافته و... محاسبه شود، قضاوت ردی و قبولی افراد بر اساس آن دقیق‌تر خواهد بود. در این مطالعه تفسیر تمام شاخص‌ها بر اساس نظریه تست کلاسیک و مبتنی بر شاخص‌های مرتبط با آزمون بود نه بر اساس توانمندی‌های خاص فرد آزمون دهنده که در تئوری سوال - جواب<sup>۳</sup> و تئوری تعمیم پذیری<sup>۴</sup> مورد توجه قرار گرفته است (۳۶).

### نتیجه‌گیری

ضرایب تحلیل پس‌آزمون آسکی در نوبت‌های مختلف اجرا می‌تواند به دلیل اثرات آموزشی اجرای آزمون بر یادگیری دانشجویان و بهبود کیفیت طراحی ایستگاه‌ها تغییر کند. محاسبه شاخص‌های سایکومتریک متناسب با آزمون‌های مبتنی بر عملکرد برای تحلیل ایستگاه‌ها و نیز تحلیل کل آزمون قابل اجرا است و اطلاعات مفیدی در مورد نقاط مثبت و منفی آزمون در اختیار می‌گذارد که برای ارائه راهکارهای اصلاحی حائز اهمیت است.

**تشکر و قدردانی:** محققان بدین وسیله مراتب تقدیر و تشکر خود را از کلیه کارکنان مرکز مهارت‌های بالینی دانشکده پزشکی که در جمع‌آوری اطلاعات نهایت همکاری داشتند، اعلام می‌دارند.

قبول داشته باشند و بدین ترتیب ضریب دشواری این نوع از آزمون‌ها به عدد یک نزدیک باشد (۳۴). در مطالعه ما در ایستگاه‌های مختلف آزمون از نوبت اول به نوبت چهارم، ضریب دشواری ایستگاه‌ها به سمت یک نزدیک‌تر می‌شود. این موضوع می‌تواند به دلیل اثرات آموزشی حاصل از آزمون برای دانشجویان، ارائه بازخورد به گروه‌های آموزشی برای آموزش مواردی که دانشجویان در آن حوزه‌ها ضعف عملکردی دارند و همچنین طراحی اثربخش‌تر ایستگاه‌ها باشد.

همچنین در مطالعه ما، ضریب همبستگی پیرسون برای دو متغیر پیوسته معدل کل و نمره نهایی آسکی دانشجویان نشان‌دهنده ارتباط مثبت متوسط بود. بدین ترتیب دانشجویانی که معدل کلی بالاتری طی دوره آموزشی کسب کرده بودند، در آزمون آسکی انتهای دوره پزشکی نیز از نمره کل بالاتری برخوردار بودند که با سایر مطالعات هم‌راستا است (۳۵). با این حال، معدل کل دانشجویان، تحت تأثیر عوامل مختلفی قرار دارد و ممکن است متغیر خوبی برای پیش‌بینی نمره آزمون صلاحیت بالینی دانشجویان نباشد. به علاوه، معدل کل دانشجویان عمدتاً متمرکز بر ارزیابی عملکرد دانشجویان در آزمون‌های تئوری است، در حالی که آزمون صلاحیت بالینی یک آزمون مبتنی بر عملکرد بالینی است. بنابراین می‌توان انتظار داشت که همبستگی قوی بین این دو آزمون بالا نباشد. به این منظور بررسی همبستگی آزمون صلاحیت بالینی با سایر آزمون‌های مشابه بالینی، بتواند معرف همبستگی بین نمرات دانشجویان در دو آزمون باشد.

از جمله نقاط قوت مطالعه، بررسی روند تغییرات شاخص‌های مورد نظر طی ۴ دوره برگزاری متوالی آزمون‌های

### REFERENCES

1. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Medical education*. 2004;38(10):1111-6.
2. 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(10):802-11.
3. NOUHI E, Motesadi M, HAGHDOUST A. CLINICAL TEACHERS'VIEWPOINTS TOWARDS OBJECTIVE STRUCTURED CLINICAL EXAMINATION IN KERMAN UNIVERSITY OF MEDICAL SCIENCE. 2008.
4. Mavis BE, Henry RC. Between a rock and a hard place: finding a place for the OSCE in medical education. *Medical education*. 2002;36(5):408-9.
5. Bakhsh T, Sibiany A, Al-Mashat F, Meccawy A, Al-Thubaity F. Comparison of students' performance in the traditional oral clinical examination and the objective structured clinical examination. *The Saudi Medical Journal*. 2009;30 (4).
6. Majumder MAA, Kumar A, Krishnamurthy K, Ojeh N, Adams OP, Sa B. An evaluative study of objective structured clinical examination (OSCE): students and examiners perspectives. *Advances in Medical Education and Practice*. 2019;10:387.

# Investigating the Trend of Changes in the Quality of Objective Structured Clinical Examination to Assess the Clinical Competence of Medical Students in Tehran University of Medical Sciences

## Zeinab Ghaffari

Assistant Internal Medicine, Shahid Beheshti University of medical sciences, Tehran, Iran

## Mahboobeh Khabaz Mafinejad, MD

Associate Professor, Department of Medical Education, Health Professions Education Research Center, Education Development Center (EDC), Tehran University of Medical Sciences, Tehran, Iran

## Fatemeh Shahbazi

Instructor of Medical Education, Department of Community Medicine, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran.

## Arash Okazi

Department of Forensic Medicine, Tehran University of Medical Sciences, Tehran, Iran

## Ehsan Bastanhigh, MD

Associate Professor, Department of Anesthesiology, Tehran University of Medical Sciences, Tehran, Iran

## ABSTRACT

**Introduction:** Considering that in recent years the clinical competency test is performed to measure the performance of medical students before graduation, it is necessary to study its psychometric properties in order to make managerial decisions to hold a valid and reliable test.

**Materials and Methods:** In this retrospective study, all medical students of Tehran University of Medical Sciences who participated in the clinical competency test during the academic year 2017-2018 were included in the study. Participants' scores and the total score of the OSCE test were collected and analyzed separately for each station. In order to evaluate the quality of OSCE, Difficulty Index, Discrimination Index,  $R^2$  and differentiation of degrees were calculated using SPSS software. The reliability of the test was determined using Cronbach's alpha. The correlation between the total grade point average of the students was calculated by the final OSCE score.

**Results:** 772 interns were included in the study in the form of 4 courses of competency test. The value of  $R^2$  for each test is 0.33 to 0.78%, 0.75 to 0.91%, 0.63 to 0.90%, 0.53 to 0.89%, respectively. The Cronbach's alpha test was obtained from 0.52 to 0.76. The range of Discrimination Index for each test was equal to -0.02 to 0.52, -0.09 to 0.64, -0.05 to 0.64, and 0.94 to 1 respectively. The range of Difficulty Index for each test was 0.68 to 0.93 and 0.21 to 0.88, 0.62 to 0.91, and 0.72 to 0.96, respectively. The correlation coefficient between the total grade point average and the final OSCE score was calculated to be 0.56 ( $P < 0.001$ ).

**Conclusion:** The internal consistency of the tests was acceptable. The R-index varied at different test times at different stations. The range of Difficulty Index and Discrimination Index in different test times have changes that can be resulting from the educational effects of test implementation for students and improving the design quality of stations. There was a direct and positive relationship between students' GPA and OSCE final score based on Pearson correlation coefficient.

**Keywords:** competence, OSCE, medicine, psychometric

## Introduction

Performance evaluation of medical students is always considered as one of the most sensitive tasks facing educational institutions (1). For this purpose, medical

education specialists have been searching for reliable methods to effectively evaluate the knowledge and clinical efficiency of students (2). A review of the evidence shows that despite the many methods of clinical evaluation, usually the evaluation of students