

بررسی روایی و پایایی آزمون OSCE در ارزیابی مهارت‌های بالینی دانشجویان سال چهارم پرستاری دانشکده پرستاری و مامایی حضرت فاطمه (س) شیراز دکتر مرضیه معطری، شادی عبدالله زرگر، دکتر مسعود موسوی نسب، دکتر نجف زارع، پروین بیگی مروست *

* دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی شیراز

چکیده

سابقه و هدف: امتحان بالینی با ساختار عینی یکی از بهترین روش‌های سنجش ابعاد شناختی، عاطفی و روان حرکتی دانشجویان پرستاری است. پژوهش حاضر به منظور بررسی روایی و پایایی آزمون OSCE (Objective Structured Clinical Evaluation) بر روی دانشجویان سال چهارم پرستاری انجام شد.

روش بررسی: در این مطالعه توصیفی از نوع همبستگی، طبق اولویت‌بندی کارشناسانه از بین کلیه مهارت‌های پرستاری، ۱۰ مهارت انتخاب و فهرست و آرس برای عملکردهای منتخب تدوین شد. ۳۷ دانشجویی طی دو روز متوالی در ۱۰ ایستگاه OSCE این مهارت‌ها را انجام دادند و در هر ایستگاه دو مشاهده‌گر، مطابق فهرست‌های و آرسی، آنها را ارزیابی کردند. روایی ملاکی (همبستگی میانگین‌های نمرات دروس بالینی و نظری پرستاری به تفکیک با نمره به‌دست آمده از آزمون OSCE) و همسانی درونی (همبستگی نمرات کل دانشجویان در OSCE با نمرات دانشجویان در هر ایستگاه) محاسبه شد. همچنین پایایی ارزیابان (همبستگی بین نمرات گزارش شده توسط دو مشاهده‌گر در هر ایستگاه به‌طور جداگانه) و دو نیمه‌سازی آزمون (همبستگی نمرات ایستگاه‌های فرد و زوج) محاسبه گردید. یافته‌ها: ضریب همبستگی نمرات OSCE با میانگین نمرات نظری و بالینی به ترتیب $0/38$ ($p = 0/031$) و $0/52$ ($p = 0/005$) بود. در همسانی درونی، بالاترین ضریب همبستگی $0/77$ بود. کمترین و بیشترین مقدار ضریب همبستگی در پایایی ارزیابان به ترتیب $0/38$ و $0/95$ بود. در دو نیمه‌سازی آزمون نیز ضریب همبستگی دو نیمه آزمون $0/61$ ($p < 0/05$) بود.

نتیجه‌گیری: استفاده از OSCE، به‌عنوان روشی معتبر و پایا برای سنجش مهارت‌های پرستاری توصیه می‌شود.

واژگان کلیدی: ارزشیابی بالینی، OSCE، روایی ملاکی، روایی سازه، پایایی.

مقدمه

از بهترین روش‌های سنجش، می‌تواند میزان تحقق اهداف آموزش در زمینه‌های شناختی، عاطفی و روان حرکتی دانشجویان پرستاری را مورد ارزیابی قرار دهد. OSCE (Objective Structured Clinical Evaluation)، آزمون ارزیابی صلاحیت بالینی است که به‌صورت عینی در ایستگاه‌های مختلف سازمان‌دهی می‌شود و از آزمون‌شوندگان خواسته می‌شود در هر ایستگاه وظایف بالینی خاصی را انجام دهند (۱). این آزمون اولین بار توسط هاردن (۱۹۷۵) طراحی و در مجله آموزش پزشکی ارائه شد. مدرسین مدرسه پرستاری دانشگاه مک ماستر نیز در سال ۱۹۸۴ برای اولین بار از آن برای سنجش مهارت‌های پرستاری در مراقبت‌های اولیه

اهمیت آموزش بالینی به‌عنوان اصلی‌ترین بخش آموزش پرستاری مورد پذیرش کلیه برنامه‌ریزان و مدیران برنامه‌های آموزشی است. از آنجایی که ارزشیابی بالینی در پرستاری، به‌عنوان یکی از ارکان مهم آموزش، باید شایستگی و توانمندی‌های پرستار را اندازه‌گیری کند، استفاده از روش‌های متفاوت ارزشیابی برای سنجش ابعاد مختلف عملکرد پرستاری کاملاً لازم است. امتحان بالینی با ساختار عینی به‌عنوان یکی

پس از اخذ مجوز انجام پژوهش، انتخاب مواد آزمون OSCE به منظور کسب اطمینان از روایی محتوایی آزمون در اولویت قرار گرفت و فهرستی از کلیه عملکردهای پرستاری تهیه شد. بدین منظور کتب مرجع مهارت‌های پایه پرستاری (۵) و فهرست‌های واریسی جهت انجام مهارت‌های پایه پرستاری (۶) مورد مطالعه قرار گرفت و فهرستی از عملکردهای پرستاری (۶۳ عملکرد) تهیه شد. در مرحله بعد به منظور اولویت‌بندی عملکردها بر اساس نظرات کارشناسی، فهرست مذکور در اختیار کلیه مدرسین گروه داخلی و جراحی و مربیان منتخب سایر گروه‌ها قرار گرفت و از آنان درخواست شد تا بر اساس اهمیت و قابلیت اجرا به رتبه بندی عملکردها بپردازند. علاوه بر این، فهرست مذکور در اختیار سرپرستاران شایسته شاغل در بیمارستان‌ها قرار گرفت تا بازتاب نظرات آنان نیز در انتخاب مواد آزمون OSCE مورد توجه قرار گیرد. پس از جمع‌آوری فرم‌های اطلاعاتی، میانگین و انحراف معیار رتبه‌های هر یک از عملکردها تعیین شد و ۱۰ عملکردی که بالاترین میانگین رتبه‌ای را داشته و ارزشیابی آنها امکان‌پذیر بود، جهت آزمون OSCE انتخاب شدند. این موارد به ترتیب شامل نحوه انجام ساکشن تراکئوستومی، نحوه انجام احیاء قلبی-ریوی، مراقبت از تراکئوستومی، نحوه شستن دست به روش جراحی، نحوه گذاشتن سوند فولی، نحوه تزریق زیرجلدی، نحوه پوشیدن دستکش به روش جراحی، نحوه دادن غذا از راه لوله بینی-معدی، نحوه تزریق عضلانی و نحوه گذاشتن لوله بینی-معدی بود. انتخاب این موارد بر اساس نظرات کارشناسی مربیان و سرپرستاران، صرفاً در جهت تامین روایی محتوایی این آزمون بود.

بعد از انتخاب عملکردها، طی یک جلسه توجیهی فهرست‌های واریسی تهیه شده برای هر یک از عملکردها در اختیار اعضای هیات علمی گروه قرار گرفت و از آنها درخواست شد تا فهرست‌ها را مورد بازبینی قرار دهند و نظرات اصلاحی خود را ارائه دهند. فهرست واریسی مجموعه‌ای از سوالات، کلمات و یا جملاتی در رابطه با هر عملکرد بود که شخص ارزشیاب بدون اطلاع از ارزش واقعی اظهار نظر خود، جلوی سوال، عبارت و یا جمله‌ای که بیش از همه مبین خصوصیات و یا صفات مورد نظر فرد ارزشیابی شونده بود، علامت می‌گذاشت (۱). بدین ترتیب اگرچه در آزمون OSCE تعداد محدودی عملکرد مورد سنجش قرار گرفت، اما هر یک از این عملکردها در فهرست واریسی مربوط به خود از تعداد زیادی مهارت ملزوم تشکیل شده بود. مشاهده و ارزشیابی عملکردها بر عهده کسانی بود که در مورد ارزشیابی آموزش دیده بودند. در این پژوهش

دانشجویان سال سوم استفاده کردند (۲). در این آزمون شانس دخالتی ندارد، شرایط برای همه یکسان است و نمره‌گذاری آن دقیق است. همچنین از معیارهای یک آزمون عادلانه برخوردار است. معیارهای یک آزمون عادلانه شامل روایی، پایایی، عینیت و عملی بودن است. روایی به معنای سنجش اهداف آموزشی و پایایی به معنای کسب نتایج یکسان از تکرار آزمون در شرایط مشابه است (۳).

متأسفانه در حال حاضر بر نحوه برگزاری امتحانات معمول بالینی ایراداتی وارد است که از آن جمله می‌توان به تعداد زیاد شرکت کننده در هر امتحان، زمان کوتاه اختصاص یافته برای ارزشیابی و غیر عینی بودن ارزشیابی‌ها اشاره کرد (۱). اگر چه به نظر می‌رسد آزمون OSCE از اکثر این ایرادات مبری باشد، استفاده از آن در برنامه‌های آموزشی گسترش نیافته است. در ایران نیز مساله ارزشیابی بالینی همچنان جزو مشکلات لاینحل باقی مانده است و نبود استانداردهای مشخص کار در محیط بالینی و نبود اهداف مشخص و مدونی که ویژگی‌های اختصاصی بودن، قابل دستیابی بودن، قابل اندازه‌گیری بودن، مناسب بودن و زمان‌بندی شده را داشته باشد، روند یادگیری دانشجویان را تحت تاثیر قرار داده است (۴). تحقیقات فراوانی در رابطه با آزمون OSCE در نقاط مختلف جهان انجام شده است، اما طبق شواهد موجود تعداد پژوهش‌های انجام شده در زمینه ارزشیابی به روش OSCE در پرستاری ایران بسیار محدود است، به طوری که تنها یک پژوهش در این زمینه یافت شد که آن هم منحصر در ارتباط با مهارت‌های اختصاصی بخش نوزادان بود (۱). اهمیت و لزوم برگزاری آزمون معتبر برای ارزشیابی مهارت‌های بالینی پرستاری و کمبود پژوهش در این زمینه موجب شد تا پژوهش حاضر به منظور بررسی روایی و پایایی آزمون OSCE در ارزیابی مهارت‌های بالینی دانشجویان سال چهارم پرستاری به اجرا در آید.

مواد و روشها

این پژوهش یک مطالعه توصیفی از نوع همبستگی است که در آن به ارزشیابی آزمون OSCE می‌پردازد. ۳۷ پرستار داوطلبی که در نیمه دوم سال چهارم پرستاری مشغول به تحصیل بودند و در حال رسیدن به مرحله دانش‌آموختگی بودند، در این پژوهش شرکت کردند. علت انتخاب این گروه به دست آوردن یک برآورد کلی از وضعیت عملکردی دانش‌آموختگان بود.

تعداد موارد موجود در فهرست‌های واریسی بین ۱۴ تا ۵۱ مورد و در کل ۲۳۴ مورد بود.

در مرحله بعد کلیه دانشجویان مورد پژوهش طی یک جلسه توجیهی که با حضور تعدادی از مسوولین دانشکده پرستاری برگزار شد، با اهداف پژوهش و نحوه برگزاری آزمون آشنا شدند و به آنها اطمینان داده شد که شرکت آنها در آزمون تاثیری در ارزشیابی آنان نداشته و اطلاعات به دست آمده فقط جهت تحلیل نتایج پژوهش مورد استفاده قرار خواهد گرفت. هدف از این جلسه جلب توجه و ترغیب دانشجویان برای شرکت در پژوهش بود.

آزمون OSCE در دو روز متوالی انجام شد. در هر یک از دو روز، ۵ ایستگاه برای ارزشیابی دانشجویان طراحی شد. جهت پیشگیری از تبادله نظر بین دانشجویان ابتدا از آنها خواسته شد تا در کلاسی که به همین منظور در نظر گرفته شده بود، حضور یابند و شماره رمز خود را به عنوان شماره هویت خود در آزمون دریافت نمایند. دانشجویان به ترتیب شماره‌های رمز وارد محیط پژوهش می‌شدند و در هر ایستگاه پس از مطالعه راهنمای آن ایستگاه به انجام عملکرد مورد نظر می‌پرداختند. در طی انجام عملکرد، دو مشاهده‌گر که فهرست واریسی را در اختیار داشتند، آنها را مورد ارزیابی قرار می‌دادند. هر دانشجویی پس از پایان ایستگاه آخر محل آزمون را ترک می‌کرد. پس از عبور نفر اول از ایستگاه اول، نفر دوم جهت شرکت در آزمون به ایستگاه اول فرا خوانده می‌شد. کلیه دانشجویان به همین ترتیب در آزمون OSCE شرکت نمودند. در طی آزمون هیچ‌یک از دانشجویان شرکت‌کننده در آزمون امکان ملاقات با دانشجویان منتظر برای فراخوانی به امتحان را نداشتند. به منظور ایجاد سرگرمی برای دانشجویانی که در انتظار فراخوانی بودند، فیلم ویدئویی پخش گردید. با توجه به اینکه مدت زمان لازم برای انجام عملکردها بین ۵ تا ۷ دقیقه متغیر بود، کلیه عملکردهایی که برای آن ۵ دقیقه وقت در نظر گرفته شده بود، در روز اول و کلیه عملکردهایی که برای آن ۷ دقیقه وقت در نظر گرفته شده بود، در روز دوم مورد ارزیابی قرار گرفت.

جهت تعیین روایی ملاکی آزمون به یک آزمون ملاک که بازتاب قابل قبولی از عملکردهای پرستاری دانشجویان باشد نیاز بود. بدین منظور از میانگین دروس بالینی پرستاری (۱۹ واحد) و میانگین نمرات دروس نظری پرستاری (۲۵ واحد) به عنوان ملاک تعیین روایی استفاده شد. همبستگی هر یک از این میانگین‌ها به صورت جداگانه با نمره به‌دست آمده از ۱۰ ایستگاه OSCE محاسبه گردید. جهت تعیین همسانی درونی،

همبستگی نمرات کل دانشجویان در OSCE با نمرات دانشجویان در هر ایستگاه محاسبه شد. جهت تعیین پایایی ارزیابان، همبستگی بین نمرات گزارش شده توسط دو مشاهده‌گر در هر ایستگاه به‌طور جداگانه محاسبه گردید. در روش دو نیمه‌کردن نیز همبستگی بین نمرات ایستگاه‌های زوج و فرد آزمون محاسبه گردید.

یافته‌ها

در بررسی روایی ملاکی آزمون، ضریب همبستگی بین نمرات OSCE با میانگین نمرات نظری و میانگین نمرات بالینی به ترتیب 0.38 ($p=0.031$) و 0.52 ($p=0.005$) بود. همان‌طور که مشاهده می‌شود ارتباط معنی‌دار ضعیف تا متوسطی بین میانگین نمرات OSCE با میانگین نمرات بالینی و نظری وجود دارد.

جهت بررسی همسانی درونی، همبستگی میان نمره کل دانشجویان در OSCE با میانگین نمره دانشجویان در هر یک از ایستگاه‌ها بررسی شد. ایستگاه تغذیه از راه لوله بینی-معدی از بالاترین ضریب همبستگی (0.74) و ایستگاه پوشیدن دستکش به روش جراحی از پایین‌ترین ضریب همبستگی (0.05) برخوردار بودند.

جهت تعیین پایایی ارزیابان، همبستگی بین نمرات مشاهده‌گران در هر ایستگاه مورد محاسبه قرار گرفت (جدول ۱). ضرایب همبستگی بین نمرات مشاهده‌گران در کلیه ایستگاه‌ها (به جز در یک مورد) از 0.7 بیشتر بود. با این وجود رابطه بین نمرات مشاهده‌گران در تمامی ایستگاه‌ها معنی‌دار بود.

جهت تعیین پایایی دو نیمه آزمون، همبستگی بین میانگین نمرات حاصل از ایستگاه‌های فرد با میانگین نمرات کسب شده از ایستگاه‌های زوج و نیز همبستگی بین میانگین دو نیمه آزمون پس از حذف نمرات ایستگاه نهم دانشجویان (به دلیل عدم همبستگی با نمره کل آزمون) بررسی شد. میانگین نمرات دو نیمه آزمون به ترتیب $12/13$ و $11/36$ و ضریب همبستگی بین این دو نیمه آزمون $0/61$ ($p < 0/001$) بود. هم‌چنین ضریب همبستگی میان میانگین نمرات نیمه اول آزمون (ایستگاه‌های فرد) پس از حذف نمرات ایستگاه نهم یعنی پوشیدن دستکش به روش جراحی با میانگین نمرات نیمه دوم آزمون (ایستگاه‌های زوج) برابر $0/65$ ($p < 0/001$) بود.

جدول ۱- ضریب همبستگی بین نمرات ارزیابان در هر

P - Value	ایستگاه	
	OSCE های	ضریب همبستگی
	نمرات هر دو ارزیاب	
< ۰/۰۰۱	۰/۸۲	ساکشن تراکتوستومی
< ۰/۰۰۱	۰/۸۸۲	تعویض پانسمان تراکتوستومی
< ۰/۰۰۱	۰/۷۳	گذاشتن سوند فولی
< ۰/۰۰۱	۰/۹۵	احیاء قلبی-ریوی
< ۰/۰۰۱	۰/۸۷	شستن دست به روش جراحی
< ۰/۰۰۱	۰/۸۱	گذاشتن لوله بینی-معدی
< ۰/۰۰۱	۰/۹۰	تغذیه از راه لوله بینی-معدی
< ۰/۰۰۱	۰/۸۰	تزریق عضلانی
۰/۰۲۱	۰/۳۸	پوشیدن دستکش به روش جراحی
< ۰/۰۰۱	۰/۷۷	تزریق زبر جلدی
< ۰/۰۰۱	۰/۹۶	نمره کل

بحث

این پژوهش نشان داد که بین نمرات آزمون OSCE با نمرات دروس نظری پرستاری ارتباط مثبت ضعیف اما معنی‌داری وجود دارد. همچنین بین نمرات آزمون OSCE با نمرات دروس بالینی پرستاری نیز ارتباط مثبت و معنی‌داری وجود دارد. با توجه به این نتایج می‌توان روایی ملاکی قابل قبول OSCE را در این پژوهش تایید کرد. برای داده‌های ملاکی چهار ویژگی مطلوب ذکر شده است که این ویژگی‌ها به ترتیب اهمیت، ربط داشتن، بی‌طرفی، پایایی و در دسترس بودن هستند. بنابراین منطقی است که بپذیریم میانگین نمرات دروس بالینی و نظری از تمامی این ویژگی‌ها برخوردارند (۷). زیرا میانگین نمرات ۲۵ واحد دروس نظری و ۱۹ واحد دروس عملی، حاصل نمرات به‌دست آمده از دروس و بخش‌های بالینی متفاوت و تحت نظارت مربیان متفاوت با شرایط آموزشی و یادگیری متفاوت می‌باشد.

اما به هر حال وجود شکاف بین دروس نظری و عملی همواره به عنوان چالشی فراروی صاحب‌نظران امر آموزش قرار دارد. وجود همبستگی ضعیف بین نتایج آزمون OSCE با دروس نظری مؤید این شکاف است. با در نظر گرفتن این نکته که آزمون OSCE باید هر سه حیطه دانش، مهارت و نگرش را مورد ارزشیابی قرار دهد، پیدا کردن ملاکی که هر سه حیطه به‌ویژه حیطه نگرشی را به طور مستقیم پوشش دهد، عملاً ممکن نیست. اما Jensen و همکارانش ذکر می‌کنند که آزمون‌های کتبی به‌خوبی می‌توانند فارغ‌التحصیلان را از نظر سطح شایستگی جداسازی کنند (۸). همچنین معطری و

همکاران ادعا می‌کنند شواهدی حمایتی برای این پیش‌فرض که بین دانش و عملکرد در مهارت‌های بالینی همبستگی وجود دارد، فراهم می‌باشد (۹).

سینگلتون و همکارانش در تحقیق خود جهت بررسی اعتبار OSCE از نظرات دانشجویان پزشکی و ارزیابان استفاده نمودند که نتایج حاصل بیانگر اعتبار خوب OSCE بود (۱۰). با توجه به نتایج حاصله و ضریب همبستگی قابل قبولی که بین نمرات OSCE با نمرات هر یک از ایستگاه‌ها به دست آمد، می‌توان نتیجه گرفت که ایستگاه‌های طراحی شده در این پژوهش همگی (بجز یک مورد) یک حوزه رفتاری را اندازه‌گیری می‌کنند و همسانی درونی آزمون در حد مطلوبی قرار دارد. در این رابطه احتمال وجود اشکال در دستورالعمل ایستگاه، طراحی ایستگاه و یا ارزیابان این ایستگاه را نمی‌توان مردود دانست. عزیزی اعتبار ارزشیابی به روش OSCE را در حد بسیار خوب ارزیابی کرده است (۱۱). ویل‌کین‌سون نیز اعتبار سازه OSCE را با استفاده از تعیین ارتباط نمرات ایستگاه‌ها با نمرات کل آزمون OSCE مورد تایید قرار داده است (۱۲).

جهت تعیین پایایی آزمون OSCE، ضریب همبستگی بین نمرات داده شده توسط مشاهده‌گران اول و دوم در تمامی ایستگاه‌ها محاسبه شد. ضرایب همبستگی بین نمرات دو مشاهده‌گر در ایستگاه‌های مختلف بین ۰/۳۸ تا ۰/۹۵ بود و همه از نظر آماری معنی‌دار بودند. نتایج همبستگی بین دو نیمه آزمون نیز مثبت و معنی‌دار بود. با توجه به اینکه ضریب همبستگی میان دو مشاهده‌گر در ایستگاه پوشیدن دستکش به روش جراحی پایین بود (۰/۳۸) و همبستگی نمرات همین ایستگاه با نمره کل آزمون نیز ضعیف و غیر معنی‌دار بود، نمرات مربوط به آن در محاسبه مجدد پایایی در نظر گرفته نشد. پس از حذف این نمرات، مجدداً ضریب همبستگی بین دو نیمه آزمون در این مرحله به ۰/۶۵ ($p < ۰/۰۰۱$) افزایش یافت. با این نتیجه، پایایی آزمون به روش دو نیمه کردن با اطمینان بیشتری مورد تایید قرار گرفت و نشان داد آزمون برگزار شده از پایایی مطلوبی برخوردار است. در این مورد عزیزی و همکارانش نیز به نتایج مشابهی دست یافتند. در پژوهش آنها پایایی آزمون OSCE ۰/۶۳ گزارش شد و پایایی نمره‌گذاری بین ارزیابان بین ۰/۶۶ تا ۰/۹۹ بود. هودگز و همکارانش نیز در تحقیقی که پیرامون ارزشیابی مهارت‌های ارتباطی بوسیله OSCE انجام دادند، به پایایی ۰/۵۹ تا ۰/۶۳ رسیدند (۱۰). لازم به ذکر است استفاده از سایر روش‌های تعیین پایایی، نظیر آزمون-بازآزمایی به علت نیاز به طراحی،

دانشجویان و همین‌طور کشف نقاط ضعف و قوت برنامه‌ریزی دانشکده‌ها توسط اساتید و مسوولین اشاره کرد. آزمون OSCE، هم‌چنین در حیطه درمان و امور بالینی روش مناسبی برای ارزشیابی پرستاران تازه‌کار و یا ارزشیابی عملکردهای جدید و بخشی از ارزیابی برنامه آموزش مداوم است.

صرف وقت و هزینه زیاد در رابطه با آزمون OSCE کاربرد چندانی ندارد. در مجموع برای پاسخ‌گویی به سوالات پژوهش می‌توان روایی و پایایی آزمون OSCE برگزار شده در دانشکده پرستاری و مامایی شیراز را مورد تایید قرار داد. استفاده از OSCE، هم در آموزش و هم در درمان، کاربردهای فراوانی دارد که از جمله این کاربردها در حیطه آموزش می‌توان به خود ارزیابی، ایجاد انگیزه و تقویت ابعاد مختلف یادگیری در

REFERENCES

- ۱- زارعی دهباغی ف. بررسی کیفیت ارزشیابی به روش OSCE در آموزش بالینی دانشجویان پرستاری بخش نوزادان. پایان نامه کارشناسی ارشد دانشگاه علوم پزشکی شهید بهشتی تهران، دانشکده پرستاری. ۱۳۸۰.
2. Ross M. Using the OSCE to measure clinical skills performance in nursing. *J Advanc Nurs* 1988; 13: 45-56.
۳. رضوی م. امتحانات بالینی سازمان دهی شده، از جزوات مرکز توسعه آموزش دانشگاه تهران. ۱۳۸۰.
۴. معطری م، پارسا یکتا ز. ارزشیابی بالینی (عملکرد): ارائه یک مدل. پروژه درس اصول و روشهای ارزشیابی در برنامه دکترای پرستاری. دانشگاه علوم پزشکی تبریز. ۱۳۷۸.
5. Potter P, Perry AG. *Fundamentals of nursing*. 5th ed. St louis: Mosby Co. 2001.
6. Te K, Delaune SC, Ladner PK. *Skills checklist to accompany fundamentals of nursing standard and practice*. Second ed. Australia: Delmar Co. 2002.
۷. سیف ع. روشهای اندازه گیری و ارزشیابی آموزشی. تهران: نشر دوران، ۱۳۸۰.
8. Jenson JJM, Scherpbier AS, Metz Jc, Grol RP, Vander veluten CP, Rethans JJ. Performance based assessment in continuing medical education for general practitioners. *J Med Educ* 1996; 130: 339-44.
۹. معطری م. ارزشیابی برنامه مدون آموزش مداوم پزشکان عمومی شهر شیراز. پایان نامه کارشناسی ارشد آموزش پزشکی، مرکز مطالعات و توسعه پزشکی دانشگاه علوم پزشکی شهید بهشتی تهران، ۱۳۷۸.
10. Singgelton A, Smith F, Harvis T, Ross-Harper R, Hilton S. An evaluation of team objective structured clinical examination (TOSCE). *J Med Educ* 1999; 33: 34-41.
۱۱. عزیزی ف. روشهای یادگیری و تحقیق در علوم پزشکی. تهران: انتشارات جوان، ۱۳۷۱.
12. Wilkinson TJ, Newble DL, Wilson PD, Carter JM, Helms RM. Development of a three-center simultaneous objective structured clinical examination. *Med Educ* 2000; 34: 798-807.
13. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. *J Med Educ* 1996; 30: 38-43.