مقاله مروری

مروری بر نقش، جایگاه و کاربرد اندازه‌گیری سایکومتریک در آزمون‌های آموزش پزشکی

مریم اکبری لاکه، سارا شهبازی*، نسرین خواجه علی

چکیده

مقدمه: اندازه‌گیری در آموزش، جایگاه ویژه‌ای دارد. تلاش برای به دست آوردن مقادیر عددی برای صلاحیت‌های مورد اندازه‌گیری که بتوان از آنها به عنوان مبنایی برای تصمیم‌گیری استفاده نمود، چیزی است که دقیقا چالش امروز اندازه‌گیری‌های سایکومتریک است. لذا هدف از این مطالعه مرور بر نقش، جایگاه و کاربرد اندازه‌گیری سایکومتریک در آزمون‌های آموزش پزشکی است.

روش‌ها: این مطالعه مروری نقلی با استفاده از کلیدواژه‌های Measurement، Assessment، Psychometric، Theoretical support، Theoretical Analysis، Theoretical foundation، Psychometric Analysis، Test، Examination، Medical Education، Validity و Reliability در سایت‌های مختلف Google Scholar، Pub Med، SID، ERIC و Science direct بدون توجه به محدودیت زمانی انجام شد. در جستجوی اولیه در مجموع ۴۲۸ مقاله پیدا شد که در بررسی اولیه عناوین مقالات، ۳۱۰ مقاله حذف و ۱۱۸ چکیده مقاله مورد بررسی قرار گرفتند و از این تعداد ۹۲ مقاله مورد مطالعه کامل قرار گرفت.

نتایج: از مجموع ۹۲ مقاله بررسی شده مستندات ۳۲ مقاله مرتبط و معتبر استخراج شد. که از این تعداد ۳۱ مقاله انگلیسی و یک مقاله فارسی بود. یافته‌های مهم این مطالعه اشاره به روایی، پایایی، تأثیر آموزشی، مقبولیت، هزینه و قابلیت اجرا و حساسیت آزمون دارد که از ویژگی‌های مهم در سایکومتری آزمون‌های علوم پزشکی است و نشان دهنده اهمیت ویژگی‌های سایکومتریک ارزیابی‌ها در آموزش پزشکی است.

نتیجه‌گیری: از نتایج به دست آمده چنین بر می‌آید که با توجه به اهمیت ویژگی‌های سایکومتریک در آزمون‌های آموزش پزشکی، مخصوصاً در آزمون‌های دارای حساسیت بالا، آشنایی با این ویژگی‌ها و شیوه‌های استانداردسازی آزمون‌ها، از ضروریات مورد نیاز اساتید علوم پزشکی است و به کارگیری تکنیک‌های افزایش دهنده روایی و پایایی آزمون‌ها و رعایت اصول اولیه در طراحی، اجرا و تحلیل آزمون‌ها، به سیاست‌گذاران و برنامه‌ریزان آزمون‌ها، اساتید و متولیان آموزش عالی، پیشنهاد می‌گردد.

واژه‌های کلیدی: اندازه‌گیری، آزمون، آموزش پزشکی، ارزیابی، روایی و پایایی.

مجله ایرانی آموزش در علوم پزشکی/ فروردین ۱۴۰۰؛ ۲۱(۵): ۴۳ تا ۵۴

مقدمه

آموزش پزشکی به عنوان هنر و علم یادگیری و تدریس پزشکی است(۱). یکی از زیرشاخه‌های اصلی آموزش، اندازه‌گیری است که به تعیین وضعیت اکتساب دانش،

* **نویسنده مسؤول:** سارا شهبازی، دانشجوی دکترای تخصصی آموزش پزشکی، دانشکده مدیریت و آموزش پزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.
sara.shahbazi@sbmu.ac.ir
دکتر مریم اکبری لاکه (استادیار) آموزش پزشکی، گروه آموزش پزشکی، مرکز تحقیقات آموزش علوم پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران.
(akbari_maryam59@yahoo.com)؛دکتر نسرین خواجه علی (استادیار)، آموزش پزشکی، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران.
(nasrinkh2009@yahoo.com)

تاریخ دریافت مقاله: ۹۹/۱/۱۹، تاریخ اصلاحیه: ۹۹/۲/۱۶، تاریخ پذیرش: ۹۹/۴/۲۳

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

مهارت و صلاحیت‌های تعریف شده، توسط دانشجویان می‌پردازد و راهی برای تعیین اثربخشی روش‌های آموزشی و تدریس، مواد آموزشی و عوامل مؤثر دیگر بر وضعیت آموزش و یادگیری است(۲).

اندازه‌گیری یک فرایند است که تعیین می‌کند یک شخص یا یک شیء، چه مقدار از یک ویژگی برخوردار است. این فرایند، نیاز به ابزار دارد. آزمون یک ابزار و روشی نظام‌مند برای اندازه‌گیری نمونه‌ای از رفتار است(۳و٤).

دلیل اهمیت اندازه‌گیری در بحث‌های آموزشی، جایگاه ویژه آن در تصمیم‌گیری‌های مهم است و در این زمینه، احتمالاً اولین و مهم‌ترین مفهوم، تلاش برای به دست آوردن مقادیر عددی برای صلاحیت‌های مورد اندازه‌گیری است تا بتوان از آنها به عنوان مبنایی برای تصمیم‌گیری استفاده نمود. چیزی که دقیقا چالش امروز اندازه‌گیری‌های سایکومتریک است(۵).

با توجه به پیچیدگی‌های فزاینده آموزش پزشکی مدرن، ارزیابی‌ها باید بتوانند دانش، مهارت و نگرش‌های مورد انتظار در یک فرد ارائه دهنده خدمات پزشکی را به خوبی نشان دهند(٦). تغییرات حرفه پزشکی، پیشرفت در تکنولوژی و افزایش سریع وسعت و عمق دانش پزشکی، موضوعاتی هستند که به طور مستقیم و غیر مستقیم منجر به تغییر در ارزیابی‌ها نیز، شده اند. امروزه ارزیابی‌ها به طور وسیعی جهت انتخاب افراد و برنامه‌های آموزشی و ارائه گواهی و مدارک معتبر به کار می‌روند. بنابراین استفاده از آزمون‌های با کیفیت بالا، در سراسر زنجیره آموزش پزشکی، بسیار مهم است(۷).

در اصل انتظاری که از اندازه‌گیری، خصوصاً در حیطه آموزش پزشکی می‌رود، مسأله‌ای است که چالش بزرگ متولیان آموزش علوم پزشکی است. چرا که مسأله دشوار برای اساتید پزشکی، دانستن این موضوع است که " ارزیابی با کیفیت"چیست؟(٦تا۹) لذا، علی‌رغم اهمیتی که تهیه و طراحی آزمون‌ها در علوم پزشکی دارند و نقش مهمی که در تعیین صلاحیت فارغ‌التحصیلان هستند، در بررسی‌ها و مطالعات انجام شده داخلی و خارجی، به مجموعه یکپارچه و هدف‌مندی از اهمیت اندازه‌گیری در آزمون‌های آموزش پزشکی دست نیافتیم، لذا این مطالعه با هدف، مروری بر نقش، جایگاه و کاربرد اندازه‌گیری سایکومتریک در آزمون‌های آموزش پزشکی انجام شد.

روش‌ها

این مطالعه، بررسی مروری نقلی (Narrative Review) است که در دی ماه سال ۱۳۹۸، انجام شده است و هدف از آن بررسی نقش، جایگاه و کاربرد اندازه‌گیری‌های سایکومتریک در آزمون‌های آموزش پزشکی، است. این مطالعه در پی یافتن ویژگی‌های یک آزمون با کیفیت و راه‌کارهای ارتقای ویژگی‌های سایکومتریک در آزمون‌های مورد استفاده در علوم پزشکی است.

نتایج مطالعه حاضر بر اساس مقالاتی است که در مجلات ملی و بین‌المللی منتشر شده اند و با استفاده از کلیدواژه‌های Measurement، Assessment، Psychometric، Theoretical support، Theoretical Analysis، Theoretical foundation، Psychometric Analysis، Test، Examination، Medical Education، Validity و Reliability در سایت‌های مختلف Google Scholar، Pub Med، SID، ERIC و Science direct بدون توجه به محدودیت زمانی انجام شده است.

معیارهای ورود به مطالعه عبارت بودند از وجود واژه‌های کلیدی در عنوان و چکیده مقاله و معیارهای خروج از مطالعه، نامرتبط بودن محتوای مقالات، با هدف این مقاله بود. در جستجوی اولیه در مجموع ٤۲۸ مقاله پیدا شد که در بررسی اولیه عناوین مقالات، ۳۱۰ مقاله حذف و ۱۱۸ چکیده مقاله مورد بررسی قرار گرفت و از این تعداد ۹۲ مقاله مورد مطالعه کامل قرار گرفت و در نهایت ۳۲ مقاله که دارای بیش‌ترین ارتباط و همبستگی با عنوان و هدف مقاله بود، به دقت مورد بررسی قرار گرفتند (تصویر ۱).

[Downloaded from ijme.mui.ac.ir on 2026-11-07]



تصویر ۱: فرایند انتخاب مقالات

نتایج

یافته‌های مطالعه بر اساس اهداف پژوهش دسته بندی شد و در قالب توضیحات نوشتاری و به صورت ٤ جدول ارائه شده است. از مجموع ۹۲ مقاله بررسی شده مستندات ۳۲ مقاله مرتبط و معتبر استخراج شد، که از این تعداد ۳۱ مقاله انگلیسی و یک مقاله فارسی بود. صد درصد مقالات به روایی وپایایی به عنوان ویژگی سایکومتری آزمون اشاره کرده بودند. ١٦ درصد مقالات علاوه بر روایی و پایایی به تأثیر آموزشی اشاره کرده و ۱۲درصد علاوه بر روایی و پایایی به مقبولیت، هزینه و قابلیت اجرا، حساسیت آزمون نیز اشاره داشتند. در ادامه به بررسی برخی از مقالات مرتبط با این موارد بالا پرداخته می‌شود:

۱- روایی در ارزشیابی:

کلاوزر (Clauser) به نقل از گلیسر (Glaser) اشاره می‌کند که یک آزمون ممکن است برای یک موقعیت ارزیابی از روایی بالایی برخوردار باشد اما در سایر موقعیت‌ها روا محسوب نشود. در شرایطی که هدف آزمون صرفاً ارزیابی دانش شناختی فراگیران باشد آزمون کتبی نیز یک ابزار روا به شمار می‌آید اما اگر هدف آزمون سنجش مهارت برقراری ارتباط فراگیران در برخورد با بیماران باشد در این شرایط استفاده از آزمون‌های کتبی مناسب نیست(٤)

کلبرت گتز (Colbert-Getz) نیز در مقاله خود بیان کرده است که ابزار اندازه‌گیری یک خصیصه ویژه روا باشد در حالی که برای سنجش همان خصیصه بر روی جامعه دیگر از اعتبار خوبی برخوردار نباشد، مثلا یک آزمون ممکن است برای سنجش استدلال بالینی کارآموزان پزشکی از روایی بالا برخوردار باشد ولی برای ارزیابی دستیاران مناسب نباشد(٥).

دیپاک (Deepak) نیز برخی از عوامل مؤثر بر روایی ارزیابی فراگیران راعنوان نموده است: تنوع ابزارهای ارزیابی، استفاده از بلوپرینت برای طرح سؤال، استفاده از دستورالعمل‌های طراحی سؤال، طراحی سؤالات مناسب برای ارزیابی سطوح شناختی بالا(٦). بنابراین با توجه به مقالات مختلف، یکی از شاخص‌های مهم در سایکومتری آزمون، توجه به شاخص روایی آزمون است که آیا آزمون طراحی شده آنچه که ما قصد سنجش آن را داریم را اندازه می‌گیرد یا نه.

۲- پایایی در ارزشیابی

با بررسی مقالات مختلف در حوزه سایکومتری آزمون‌های علوم پزشکی، یکی دیگر از شاخص‌های مهم در سایکومتری مفهوم پایایی است. دزجاردینز (Desjardins) در مقاله خود بیان می‌کند هرچه تفاوت بین فردی فراگیران بیش‌تر باشد (یعنی مقدار خطای تصادفی و خطای آزمون‌گران کم‌تر)، پایایی افزایش خواهد یافت. دیپاک نیز بیان می‌کند برای افزایش پایایی در آزمون لازم است محققان از ابزارهای مختلف استفاده کنند و سؤالات را همگون با اهداف دوره تنظیم نمایند تا بتوان در جهت یکی از مفاهیم سایکومتری آزمون گام برداشت. در خصوص برگزاری آزمون‌های مکرر برای افزایش پایایی آزمون نیز در مطالعات درنینگ و دیپاک (Deepak, Durning) اشاره شده است(٦و٧). کلبرت گتز

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

نیز بیان می‌کند که افزایش تعداد سؤالات و طراحی سؤالات از تمام اهداف و موضوعات می‌تواند باعث افزایش پایایی آزمون و بهبود سایکومتری آن گردد(۵). لذا برای بهبود پایایی آزمون با تنوع ابزارهای ارزیابی، برگزاری آزمون‌های مکرر، افزایش تعداد سؤالات، همگون کردن سؤالات با اهداف دوره و طراحی سؤالات از تمام اهداف و موضوعات می‌توان در جهت سایکومتری آزمون حرکت کرد.

۳- تأثیر آموزشی

درنینگ (Durning) معتقد است محتوی آزمون موجب یادگیری می‌شود. به عنوان مثال در صورتی که قصد داریم فراگیران در انتهای دوره مهارت حل مسأله کسب کنند، نباید سؤالاتی پرسیده شود که حفظیات را می‌سنجند(۸).

نکته دیگر در خصوص تأثیر آموزشی آزمون که در مقالات متعدد به آن اشاره شده است این است که طرح سؤال تشریحی در آزمون، منجر به هدایت فراگیران به یادگیری عمیق مطالب درسی می‌شود یا برگزاری آزمون آسکی باعث می‌شود دانشجویان به کسب مهارتهای عملی توجه بیش‌تر کنند(۵تا۸).

در این خصوص دزجارنیز (Desjardins) بیان می‌کند که آزمون براساس اطلاعاتی که به فراگیران می‌دهد موجب یادگیری می‌شود. شاید بتوان گفت که بازخورد دادن به دانشجو براساس نتایج آزمون ملموس ترین تأثیر مورد انتظار از برگزاری آزمون است(۷). لذا تأثیر آموزشی یکی دیگر از جنبه‌های سایکومتری آزمون است که با در نظر گرفتن آن می‌توان باعث بهبود آزمون‌ها شد.

٤- مقبولیت

یکی دیگر از جنبه‌های مهم در سایکومتری آزمون مقبولیت است. آپشتن (Epstein) معتقد است که این که ابزار تا چه حد مورد استقبال و پذیرش دست اندرکاران قرار می‌گیرد، از مسائلی است که باید به آن توجه کرد. از همین رو برای انتخاب و استفاده از ابزارهای ارزیابی، باید نظرات و عقاید دست اندرکاران را جویا شویم و از آن‌ها در مورد شیوه ارزیابی بازخورد بگیریم(۹).

۵- هزینه و قابلیت اجرا

یکی دیگر از نکات مهم در سایکومتری آزمون، هزینه وقابلیت اجرای آن است. ارزیابی خوب هزینه بر است. کلیسر معتقد است که طراحی سؤال، آموزش آزمون‌گران، اجرای آزمون، تصحیح پاسخ‌ها، تحلیل آزمون و بازخورد به دانشجویان و دست‌اندرکاران و موسسه فعالیت‌هایی هستند که مستلزم صرف هزینه، زمان و نیروی انسانی هستند(٤).

بررسی مقالات متعدد نشان می‌دهد که ارزیابی خوب با یادگیری خوب همراه است و آن را تسهیل می‌کند. این موضوع مطرح کننده نقش ارزشمندی است که ارزیابی می‌تواند در ارتقای کیفیت یادگیری فراگیران بازی کند(۸تا۱۰).

٦- حساسیت آزمون

یکی دیگر از عواملی که در سایکومتری آزمون نقش دارد حساسیت آزمون است. در مطالعات متعدد بیان شده است که آزمون‌های پیشرفت تحصیلی و آزمون‌هایی که با هدف تکوینی برگزار می‌شوند، دارای حساسیت کم هستند. این آزمون‌ها در شرایطی مانند مطالعات پایلوت و تأیید آیتم‌های جدید برای آزمون‌های دارای حساسیت بالا، بسیار حیاتی هستند(۱۰و۱۱). بنابراین، میزان تلاش آزمون شونده برای آزمون، می‌تواند بر نتیجه این آزمون‌ها تأثیرگذار باشد. این آزمون‌ها، هدایت کننده نحوه تدریس، عملکرد آینده اساتید و برنامه آموزشی هستند، لذا به دست آوردن نتایج معتبر در این آزمون‌ها حائز اهمیت است. برخی راه‌کارهای پیشنهاد شده در بررسی مقالات به دست آمده است و در جهت ارتقای ویژگی‌های سایکومتریک آزمون است در جدول ۱ آورده شده است(۱۱تا۱۸).

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

جدول ۱: برخی راه‌کارها برای ارتقای ویژگی‌های سایکومتریک

ویژگی سایکومتریک	راه‌کارهای ارتقای ویژگی سایکومتریک در آزمون
اعتبار / روایی	۱- انتخاب ابزار متناسب با اهداف مورد ارزیابی
	۲- استفاده از بلوپرینت
	۳- نمونه‌گیری کامل از کوریکولوم اجراشده
	٤- مشخص نمودن وزن نسبی هر محتوا در آزمون
	٥- آنالیز کمی نتایج پس از آزمون برای تصمیم‌گیری در مورد بانک سؤالات (ضریب دشواری، ضریب تمیز و اصلاح سؤالات بر اساس نتایج به دست آمده)
پایایی	۱- افزایش طول آزمون
	۲- افزایش تعداد سؤالات آزمون
	۳- استفاده از مصححان آموزش دیده
	٤- استفاده از تحلیل آیتم‌ها
	٥- استفاده از آزمون‌ها و سؤالات استاندارد شده
	٦- رعایت شرایط محیطی آزمون
	۷- محاسبه ضریب همبستگی بین آزمون‌های مختلف برای یک صلاحیت در یک سطح از هرم میلر
	۸- استفاده از آزمون‌هایی که ضریب پایایی بالای ۰/۸ دارند.
تأثیر آموزشی	۱- استفاده از آزمون‌های دارای روایی بالا
	۲- استفاده از آزمون‌های دارای پایایی بالا
مقبولیت	۱- استفاده از نظرات و عقاید دست اندرکاران در انتخاب و استفاده از ابزارهای ارزیابی
	۲- گرفتن بازخورد در مورد شیوه ارزیابی
هزینه و قابلیت اجرا	۱- سرمایه گذاری در بحث ارزیابی صحیح از عملکرد فراگیران
حساسیت آزمون	۱- استفاده از آزمون‌های تکوینی برای ارتقا آزمون‌های پایانی

۷- ویژگی‌های سایکومتریک در آزمون‌های علوم پزشکی

گلبرت گتز و دیپاک در مطالعات خود بیان می‌کنند که نکته مهم در آموزش پزشکی، امکان تفکیک دانشجویان دارای توانمندی بیش‌تر از افراد دارای توانمندی کم‌تر است(۵و٦). آنالیز سایکومتریک در آزمون‌های چند گزینه‌ای، امکان محاسبه ضریب دشواری، ضریب تمیز و تأثیر گزینه انحرافی را فراهم می‌کند و نقش مهمی در تهیه و بازبینی ابزار اندازه‌گیری دارد و امکان تهیه یک بانک سؤال کامل و دارای اعتبار را فراهم می‌کند(۱۱تا۱۹).

خیامی (Kheyami) و کوگان (Kogan) در مقاله خود بیان می‌کنند که امروزه سؤالات چندگزینه‌ای، در آزمون‌های ورودی و مؤسسه‌ای رشته‌های علوم پزشکی، در سراسر جهان به صورت گسترده‌ای، استفاده می‌شوند، زیرا امکان دریافت فیدبک سریع و ساختارمند در آزمون‌های آنلاین، به راحتی امکان پذیر است(۱٦و۱۷).

نامدئو (Namdeo) و لین (Lynne) معتقدند که علی‌رغم مزایای این آزمون‌ها، اشتباهات تکنیکی و عدم رعایت اصول طراحی، از اعتبار آزمون کاسته و آزمون‌گر را از اهداف خود دور می‌کند(۱۹و۲۰).

اگر این آزمون‌ها به خوبی طراحی شده باشند، توان سنجش سطوح بالاتر از دانش، مانند درک و کاربرد دانش را نیز، خواهند داشت. اما مسأله مهم این است که اگر این آزمون‌ها در موارد بالینی به کار روند، منجر به کاهش اعتبار آزمون‌ها می‌گردند، زیرا امکان تهیه بیش از ۳ گویه در یک سؤال، در موارد پزشکی به سختی ممکن است(۲۱تا۲۵). با این حال، تنها داشتن نتایج عددی برای تعیین صلاحیت‌های افراد، کافی نیست و به ابزاری نیاز است که بتوانند سطوح بالای تفکر و ترکیبی از مهارت‌هایی را که لازمه دستیابی به صلاحیت‌ها هستند را

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

مورد اندازه‌گیری قرار دهد(۵).

بنابراین، آزمون‌های آسکی تکنیک‌های شبیه سازی و آزمون‌های بالینی مانند تمرین ارزیابی بالینی کوتاه، مشاهده مستقیم مهارت‌های پروسیجرال، ارزشیابی ۳۶۰ درجه، موارد بالینی کامل و موارد بالینی کوتاه راهی برای دستیابی به این اهداف شدند(۲۶و۲۷). آزمون‌های کاربردی در آموزش پزشکی و ویژگی‌های سایکومتریک آنها که در بررسی مقالات به دست آمده، در جدول ۲، ارائه شده است.

جدول ۲: ویژگی‌های سایکومتریک در آزمون‌های علوم پزشکی

آزمون	روایی	پایایی	عینیت
شفاهی	ندارد	ندارد	ندارد
تشریحی گسترده پاسخ	ضعیف	ضعیف	ندارد
تشریحی کوتاه پاسخ	بالاتر از آزمون‌های تشریحی گسترده پاسخ	بالاتر از آزمون‌های تشریحی گسترده پاسخ	بیش از آزمون‌های تشریحی گسترده پاسخ
چند گزینه ای	دارد	دارد	دارد
جور کردنی	دارد	دارد	دارد
ویژگی‌های کلیدی	دارد	دارد	دارد (به شرط همراهی با سؤالات چند گزینه‌ای)
موارد کامل بالینی	ندارد (مگر این که ۱۰ مورد بالینی و به مدت ۲۰۰ دقیقه ارزیابی شوند)	ندارد (مگر این که ۱۰ مورد بالینی و به مدت ۲۰۰ دقیقه ارزیابی شوند)	ندارد
موارد کوتاه بالینی	روایی سازه دارد	ندارد (مگر این که تعداد موارد بالینی متعددی ارزیابی شوند)	ندارد
OSCE	دارد	دارد (به شرط دارا بودن۱۸-۱۲ ایستگاه و نیروهای آموزش دیده در هر ایستگاه)	دارد
Mini-CEX	ندارد	دارد (به شرط انجام حداقل ٦-٤ مورد)	ندارد
DOPS	ندارد	دارد (به شرط انجام حداقل ٦-٤ مورد)	ندارد
۳٦۰ درجه	دارد	تکرار پذیری دارد	ندارد
Portfolio	دارد (بسته به محتوای کارپوشه)	توافق بین ارزیابان در حدود ۰/۷-۰/٦	ندارد

۸ -نقش هرم میلر و تاکسونومی بلوم در آزمون‌ها

پیرس (Pearce) معتقد است علی‌رغم تلاش فراوان، همچنان این نگرانی وجود دارد که ارزیابی به ندرت توانسته است بر عملکرد فردی دانشجو تأثیر داشته باشد(۲۲). در اصل، مهم‌ترین مسأله، مناسب بودن یک ارزیابی است که بتواند همه جنبه‌های مختلف مد نظر یک آزمون‌گر را با معیارهای قابل قبول سایکومتریک، اندازه‌گیری نماید(۱۵). در بررسی مقالات، استفاده از هرم میلر برای انتخاب مناسب روش‌های اندازه‌گیری متناسب با اهداف انتخاب شده (جدول ۳) و تاکسونومی بلوم (جدول ٤)، برای اطمینان از پوشش کامل اهداف، توصیه می‌گردد(۱۲).

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

جدول ۳: ابزارهای ارزیابی متناسب با سطوح مختلف هرم میلر

سطح ارزیابی بر اساس هرم میلر	آزمون‌ها
می داند و می‌داند چگونه	آزمون‌های شفاهی، سؤالات تشریحی گسترده و کوتاه پاسخ، سؤالات چند گزینه‌ای، سؤالات جور کردنی، آزمون ویژگی‌های کلیدی
نشان می‌دهد چگونه	OSCE، موارد بالینی کامل و موارد بالینی کوتاه
انجام می‌دهد	تمرین ارزیابی بالینی کوتاه، مشاهده مستقیم مهارت‌های پروسیجرال، چک لیست، ارزشیابی ۳۶۰ درجه، لاگ بوک و پورتفولیو

جدول ۴: توزیع آزمون‌ها بر مبنای سطوح مختلف تاکسونومی شناختی بلوم

آزمون‌های شناختی	سطوح تاکسونومی بلوم
چند گزینه ای	یادآوری، فهمیدن، کاربرد، تحلیل
صحیح - غلط	یادآوری
جور کردنی	یادآوری و فهمیدن
تشریحی بلند پاسخ	یادآوری، فهمیدن، کاربرد، تحلیل، ارزشیابی و نوآوری
تشریحی کوتاه پاسخ	یادآوری و فهمیدن
شفاهی	یادآوری

بحث

در این مقاله به مرور و بررسی نقش، جایگاه و کاربرد اندازه‌گیری سایکومتریک در آزمون‌های آموزش پزشکی پرداخته شد. در این مطالعه نشان داده شد که اندازه‌گیری مقوله‌ای مهم در آموزش علوم پزشکی امروز است.

نتیجه مطالعه حاضر نشان داد که آزمون‌هایی با کیفیت هستند که روایی، پایایی و تأثیر آموزشی بالا داشته و ضمن قابل اجرا بودن، مقرون به صرفه هستند. در گلبرت کتز، پیرس و گایچا (Ghaicha) نیز توصیه می‌کنند بهترین آزمون‌ها، آزمون‌هایی هستند که روایی، پایایی و تأثیر آموزشی بالا داشته و ضمن قابل اجرا بودن، مقرون به صرفه باشند. ولی نکته حائز اهمیت این است که بسیاری از آزمون‌ها، مخصوصاً آزمون‌های پیشرفت تحصیلی، از روایی و پایایی قابل قبولی برخوردار نیستند(۵و۱۱و۲۲).

در همین راستا شاه (Shah) در مطالعه خود نیز به اهمیت روایی آزمون اشاره می‌کند که برای تعیین اعتبار آزمون‌های دارای حساسیت کم، شناسایی میزان تلاش آزمون شوندگان است. در این موارد می‌توان نمرات دانشجویانی را که از تلاش کم‌تری برخوردار بوده اند از تحلیل‌های سایکومتریک حذف نمود. برای این منظور استفاده از مقیاس Task Taking Effort Short، توصیه می‌گردد. بنابراین به نظر می‌رسد استفاده از تکنیک‌های معتبر سازی آزمون‌های دارای حساسیت کم، باید جدی تر تلقی گردد(۲۶).

در مقابل، استیف (Steif) و سون (Suen) در مطالعه خود بیان می‌کنند که آزمون‌های دارای حساسیت بالا، به عنوان بخشی از سیستم پاسخ‌گویی، معیاری برای تعیین کیفیت دانشگاه‌ها، ارتقای افراد به مراتب بالاتر ومبنای دریافت جایزه، عمل می‌کنند و با توجه به این که نقش عمده‌ای در سرنوشت آزمون شوندگان دارند، باید از اصول سایکومتریک صحیح برخوردار باشند(۲۷و۲۸).

توکل معتقد است برای اطمینان می‌توان با انجام یک مطالعه پایلوت، حداقل در گروهی کوچک، به کنترل ویژگی‌های سایکومتریک این آزمون‌ها پرداخته شود. ولی مسأله مهم، تضاد این کار با حفظ امنیت سؤالات است. لذا

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

می‌توان، از تحلیل داده‌های مشابه توسط گروهی از افراد خبره (Paired Item Procedure) استفاده نمود(۲۹).

سنجش اصولی دانش، مهارت و عملکرد دانشجویان پزشکی، یکی از روش‌های تضمین پاسخ‌گویی به جامعه است. از طریق آزمونی که شاخص‌های سایکومتریک شامل روایی، پایایی، اثر آموزشی، مقبولیت و قابلیت اجرا را دارد، می‌توان اطمینان حاصل کرد که پزشکان آینده، مهارت‌ها و شایستگی‌های لازم برای عملکرد خوب را کسب نموده و صلاحیت کافی دارند(۳۰تا۳۵).

در مطالعه‌ای که نصری و همکاران، بر روی دانشجویان پزشکی و در مورد امتحانات جامع علوم پایه و پیش کارورزی انجام دادند، گزارش نمودند که آزمون‌ها قادر به ارزیابی نتیجه سعی و تلاش، مهارت و نگرش دانشجو نیستند و فراگیری معلومات برای این آزمون‌ها در حل سؤالات بالینی در برخورد با بیمار بی تأثیر است. این امتحانات باعث ایجاد استرس زیاد شده و نتایج موقتی داشته و اثر آموزشی ندارد. دانشجویان نیز در نظرسنجی اعلام کردند که امتحانات با نیازهای شغلی آنها ارتباط ندارد که نشان دهنده نداشتن روایی مناسب آزمون‌های جامع علوم پایه و پیش کارورزی است(۳٤).

اسدپور و همکاران در مطالعه خود بیان کردند آزمون‌های علوم پایه پزشکی از روایی مناسب برخوردار نیستند و در نتایج مطالعه خود، به اهمیت ارتقای آموزش و آزمون‌ها پرداخته‌اند و نیاز به تجدید نظر در اجرای آنها را مطرح نمودند(۳۵). نتایج مطالعه سیاح ملی و همکاران نیزنشان داد، سطح آزمون‌های درون دانشگاهی پایین و سؤالات این آزمون‌ها در پایین ترین طبقه تاکسونومی بلوم قرار دارد(۳٦). وجود چنین مشکلی می‌تواند روایی آزمون را با مشکل مواجه سازد و موجب سوق دادن دانشجویان به دانش سطحی و محفوظات گردد. در مطالعه مروری ما نتایج نشان داد که هر چقدر یک آزمون ویژگی‌های روایی، پایایی، اثرآموزشی، مقبولیت و قابلیت اجرا را داشته باشد باعث استانداردسازی آزمون و هدایت به سمت ارزشیابی صحیح از فراگیران می‌شود.

نتایج مطالعه اکبری نشانگر این امر است که آزمون جامع علوم پایه روایی پیش بین مناسبی در رابطه با آینده تحصیلی دانشجویان دارد. لذا برگزاری آزمون‌های استاندارد از اهمیت بسیار زیادی برخوردار است(۳۷). روایی از عوامل مؤثر در سایکومتری آزمون است که در مطالعه حاضر و مطالعات گلبرت کتز، پیرس و گایچا به اهمیت آن اشاره شده است.

آنچه از نتایج مطالعات مختلف می‌توان برداشت نمود نیاز مبرم به تغییر در رویکرد ارزیابی در آموزش پزشکی است. چرا که علی‌رغم پیشرفت‌های فراوان در بحث اندازه‌گیری در آموزش پزشکی، هنوز هم چالش‌های متعددی در این زمینه وجود دارد که عبارتند از:

۱. عدم آشنایی اساتید با نحوه تهیه آزمون‌ها و اصول سایکومتریک

۲. عدم نظارت بر نحوه طراحی و تحلیل آزمون‌ها

۳. عدم آشنایی و استفاده اساتید از آزمون‌های متناسب با اهداف آموزشی

٤. عدم آشنایی و استفاده اساتید از تئوری‌های اندازه‌گیری(۵و۹و۱۰و۲۱و۲٤و۲٦).

بنابراین باید به روش‌های مختلف، شرایطی را ایجاد کرد که ویژگی‌های سایکومتریک هر آزمون، را تا سر حد امکان، تأمین نمود.

از نقاط قوت پژوهش مقاله حاضر این بود که همه عوامل مؤثر بر سایکومتری ابزار را در قالب یک چارچوب در یک مطالعه مد نظر قرار داد. با این حال این پژوهش با محدودیت‌هایی نیز روبرو بود، مرور حاضر تنها بر روی مقالات فارسی و انگلیسی زبانی انجام شد که متن کامل آنها در جستجوی اینترنتی قابل دسترسی بود ، لذا ممکن است مطالعاتی به زبان‌های دیگر موجود باشد و یا به دلیل عدم دسترسی به متن کامل آنها، در مرور حاضر مورد توجه قرار نگرفته باشند.

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

در نهایت پیشنهاد می گردد، محققان از طریق اجرای دوره‌های آموزشی آشنایی با سایکومتری ابزار به صورت تجربی، اثربخشی آزمون‌ها را در محیط‌های آموزش پزشکی کشور ارتقا داد.

نتیجه‌گیری

نتایج این مطالعه نشان‌دهنده توجه ناکافی به ویژگی‌های سایکومتریک در آزمون‌های علوم پزشکی است و این مسأله حتی در بحث پژوهش هم مورد غفلت واقع شده است. روایی، پایایی، تأثیر آموزشی، مقبولیت و دیگر ویژگی‌های سایکومتریک، اساس حفظ عدالت، اطمینان، اعتماد و پذیرش آزمون‌ها را تشکیل می‌دهند، در حالی‌که برخی از اساتید و تهیه‌کنندگان آزمون‌ها دانش قابل توجهی در این خصوص ندارند. لذا با توجه به اهمیت بسیار بالای ویژگی‌های سایکومتریک در آزمون‌های آموزش پزشکی، به خصوص در آزمون‌های دارای حساسیت بالا، استفاده از تاکسونومی بلوم، برای اطمینان از پوشش کامل اهداف و هرم میلر برای انتخاب مناسب روش‌های اندازه‌گیری متناسب با اهداف انتخاب شده، به‌کارگیری تکنیک‌های افزایش دهنده روایی و پایایی آزمون‌ها و رعایت اصول اولیه در طراحی، اجرا و تحلیل آزمون‌ها، به سیاست‌گذاران و برنامه‌ریزان آزمون‌ها، اساتید و متولیان آموزش عالی، پیشنهاد می‌گردد.

قدردانی

از کلیه اساتیدی که در فرآیند تهیه مقاله ما را راهنمایی نمودند تشکر و قدردانی می‌کنیم.

منابع

1. Blake JE. The Consequential Effects of High-Stakes Testing on Teacher Pedagogy. Practice and Identity: Teacher Voices Disrupt the A Priori [dissertation]. Knoxville, Tennessee: University of Tennessee; 2008.
2. Boulet JR, Swanson DB. Psychometric Challenges Of Using Simulations For High-Stakes Assessment. Simulations In Critical Care Education And Beyond. Des Plains, IL: Society Of Critical Care Medicine; 2004: 119-130.
3. Brookhart SM. A theoretical framework for the role of classroom assessment in motivating student effort and achievement. Applied Measurement in Education. 1997; 10(2): 161-180.
4. Clauser BE. Educational Measurement: Origins, Theories and Explications. Journal of Educational Measurement. 1998; 35(3): 273-275.
5. Colbert-Getz JM, Ryan M, Hennessey E, Lindeman B, Pitts B, Rutherford KA. Measuring Assessment Quality With An Assessment Utility Rubric For Medical Education. MedEdPORTAL. 2017. 13: 10588.
6. Deepak KK, Al-Umran KU, AI-Sheikh MH, Dkoli BV, Al-Rubaish A. Psychometrics of Multiple Choice Questions with Non-Functioning Distracters: Implications to Medical Education. Indian J Physiol Pharmacol. 2015; 59(4): 428-35.
7. Desjardins CD, Bulut O. Handbook of educational measurement and psychometrics using R. London: CRC Press; 2018.
8. Boulet JR, Durning SJ. What we measure. . . and what we should measure in medical education. Med Educ. 2019 ; 53(1): 86-94.
9. Epstein RM. Assessment in medical education. N Engl J Med. 2007 ; 356(4): 387-96.
10. Fuchs LS, Fuchs D, Hamlett CL. Effects of instrumental use of curriculum-based measurement to enhance instructional programs. Remedial & Special Education. 1989. 10(2): 43-52.
11. Ghaicha A. Theoretical Framework for Educational Assessment: A Synoptic Review. Journal of Education and Practice. 2016; 7(24): 212- 231.
12. Gunzenhauser MG. High-stakes testing and the default philosophy of education. Theory into Practice. 2003; 42(1): 51-58.
13. Haendchen Filho A, do Prado HA, Ferneda E, Nau J. An approach to evaluate adherence to the theme

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

and the argumentative structure of essays. Procedia Computer Science. 2018; 126: 788-797.
14. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, Collaborators IC. The role of assessment in competency-based medical education. Med Teach. 2010; 32(8): 676-82.
15. Huot B. Toward a new theory of writing assessment. College composition and communication. 1996; 47(4): 549-566.
16. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. Sultan Qaboos Univ Med J. 2018 ; 18(1): e68-e74.
17. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing Variable Rater Assessments As Both An Educational And Clinical Care Problem. Acad Med. 2014; 89(5): 721-7.
18. Mccoubrie P. Metrics In Medical Education. The Ulster Medical Journal. 2010; 79(2): 52-6.
19. Namdeo SK, Sahoo B. Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. Int J Res Med Sci. 2016; 4(5): 1716-1719.
20. Lynne P. Coming to Terms: A Theory of Writing Assessment. Logan, Utah: All USU Press Publications ; 2006.
21. Nolte E, Viola Fry C, Winpenny E, Brereton L. Use Of Outcome Metrics To Measure Quality In Education And Training Of Healthcare Professionals. A Scoping Review Of International Experiences; 2012. [cited 2021 feb 16]. available from: https://www. rand. org/pubs/working_papers/WR883. html
22. Pearce J, Edwards D, Fraillon J, Coates H, Canny BJ, Wilkinson D. The rationale for and use of assessment frameworks: improving assessment and reporting quality in medical education. Perspect Med Educ. 2015; 4(3): 110-8.
23. Preston R, Gratani M, Owens K, Roche P, Zimanyi M, Malau-Aduli B. Exploring the Impact of Assessment on Medical Students' Learning. Assessment & Evaluation in Higher Education. 2019; 45(2):1-16.
24. Schüttpelz-Brauns K, Kadmon M, Kiessling C, Karay Y, Gestmann M, Kämmer JE. Identifying Low Test-Taking Effort During Low-Stakes Tests With The New Test-Taking Effort Short Scale (TESS)–Development And Psychometrics. BMC Medical Education. 2018; 18(1): 101.
25. Schuwirth LW, van der Vleuten CP. How 'testing'Has Become 'programmatic Assessment for Learning'. Health Professions Education. 2018; 5(3): 177-184.
26. Shah SSH, Munir TA, Sabir M, Tipu SA. Psychometric Analysis of MCQs Used in Assessing the Students at Entrance to a Medical College. Annals of King Edward Medical University. 2012; 18(3): 296-296.
27. Steif PS, Dantzler JA. A statics concept inventory: Development and psychometric analysis. Journal of Engineering Education. 2005; 94(4): 363-371.
28. Suen HK, Wu Q. Psychometric paradox of very high-stakes assessment and solutions. KEDI Journal of Educational Policy. 2006; 3(1).
29. Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. Med Teach. 2017 ; 39(10): 1010-1015.
30. ul Islam Z, Usmani A. Psychometric analysis of Anatomy MCQs in Modular examination. Pak J Med Sci. 2017 ; 33(5): 1138–1143.
31. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. Med Educ. 2005; 39(3): 309-17.
32. Vyas R,Supe A. Multiple choice questions: a literature review on the optimal number of options. Natl Med J India. 2008; 21(3): 130-3.
33. Wrigley W, van der Vleuten CPM, Adrian Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE guide no. 71. Med Teach. 2012; 34(9): 683-97.
34. Nasri K, Kahbazi M, Nasri S. [Medical Students' Viewpoints toward Basic Sciences and Preinternship Comprehensive Exams in Arak University of Medical Sciences]. Iranian Journal of Medical Education. 2010; 10(1) :82-91. [Persian]
35. Asadpour M, Sheikh Fathollahi M, Rezaeian M, Jafari Naveh H, Mahboobi Rad M, Amir Afzali. [A Survey on Factors Affecting the Results of Comprehensive Examination of Medical Basic Sciences in 2005 to 2009 Groups of Medical Students from the Students' Viewpoints in Rafsanjan University Medical Sciences]. Journal of Rafsanjan University of Medical Sciences. 2015; 14(12): 1073-1084. [Persian]

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

36. Sayyahmelli M, Barzegar M, Bilan N, Aslanabadi S, khoshbaten M, Gasemzadeh A, et al. [Comparing Multiple-Choice Questions Quality Parameters of Pediatric, General Surgery, Internal Medicine and Genecology and Obstetrics Residency Tests in Preboard and Promotion Examinations of Tabriz University of Medical Sciences and National Board Examination in 2010 and 2011]. Journal of Medical Education Development. 2014; 8(18): 43-53. [Persian]
37. Akbari M, Sekandari S. [Evaluation of the Predictive Validity of Comprehensive Basic Science Examination for the Adequacy of Dentistry Students' Clinical Competence]. Journal of Mashhad Dentistry School. 2016; 40(2): 113-22. [Persian]

[Downloaded from ijme.mui.ac.ir on 2026-11-07]

A review of the role, status and application of psychometric measurement in medical education tests

Maryam Akbarilakeh[1], Sara Shahbazi[2], Nasrin Khajeali[3]

Abstract

Introduction: *Measurement has a special place in education. The main challenge of today psychometric measurements is to obtain numerical values for the competencies that can be used as a basis for decision making. In view of this, the purpose of this study is to review the role, position, and application of psychometric measurement in medical education exams.*

Methods: *This review study was conducted through the keywords Measurement, Assessment, Psychometric, Theoretical support, Theoretical Analysis, Theoretical foundation, Psychometric Analysis, Test, Examination, Medical Education, Validity and Reliability at websites such as Google Scholar, PubMed, SID, ERIC and Science direct and has been done on 32 articles without از time limit.*

Results: *From a total of 92 reviewed articles, the documentation of 32 relevant and valid articles was extracted with 31 English articles and a Persian article. Important findings of this study underline the validity, reliability, educational effect, acceptability, cost and applicability as well as sensitivity of the test, which are important features in the psychometrics of medical science tests. These findings highlight the importance of psychometric properties in evaluations in medical education.*

Conclusion: *The results show that due to the importance of psychometric properties in medical education exams, especially in high-sensitivity exams, familiarity with these features and standardization methods of exams is a necessity for medical educators. The use of techniques to increase the validity and reliability of exams and considering the basic principles in the design, preparation, implementation and analysis of exams is recommended to policy makers and examiners, educators and those in charge of higher education.*

Keywords: Measurement, Testing, Medical education, Assessment.

Addresses:

1. Assistant Professor , Department Of Medical Education, Center For Educational Research In Medical Sciences (Cerms), Iran University Of Medical Sciences, Tehran, Iran. email: akbari_maryam59@yahoo.com.
2. (✉) Phd Candidate, Department Of Medical Education, School Of Management & Medical Education Sciences. Shahid Beheshti University Of Medical Sciences. Tehran. Iran. email: sara.shahbazi@sbmu.ac.ir
3. Assistant Professor , Department Of Medical Education , Jundishapur University Of Medical Science, Ahvaz,Iran. email: nasrinkh2009@yahoo.com

[Downloaded from ijme.mui.ac.ir on 2026-11-07]