مقاله مروری
مجله دانشگاه علوم پزشکی رفسنجان
دوره ۱۳، اسفند ۱۳۹۳، ۱۱۷۰-۱۱۵۳

روایی و پایایی ابزارها و روش‌های مختلف اندازه‌گیری آنها در پژوهش‌های کاربردی در سلامت

ابوالفضل محمدبیگی[۱]، نرگس محمدصالحی[۲]، محمد علی گل[۳]

دریافت مقاله: ۹۳/۶/۳ ارسال مقاله به نویسنده جهت اصلاح: ۹۳/۱۱/۶ دریافت اصلاحیه از نویسنده: ۹۳/۱۱/۱۲ پذیرش مقاله: ۹۳/۱۲/۶

چکیده

امروزه با گسترش تحقیقات در علوم پزشکی و اجتماعی، پژوهش، آموزش و مدیریت، بسیاری از پدیده‌ها از طریق ابزارهایی چون پرسش‌نامه‌ها، چک لیست‌ها و تست‌های آموزشی و یا ارزیابی‌های داوران و به منظور کمی‌سازی رفتار افراد یا جامعه، مورد مطالعه قرار می‌گیرد. روایی و پایایی از شاخص‌های لازم جهت اندازه‌گیری و دستیابی به مقادیر واقعی پیامدها می‌باشد. آگاهی و فهم کافی از روایی و پایایی از اصول متدولوژی در طراحی مطالعات و کاربرد سنجش‌های مختلف است که متأسفانه در بسیاری از موارد نادیده گرفته می‌شود. هدف نوشتار حاضر آشنایی محققین و دانشگاهیان با مفاهیم پایایی، روایی و شیوه‌های مختلف اندازه‌گیری آن است.

واژه‌های کلیدی: روایی، پایایی، روان سنجی، آلفای کرونباخ، تحلیل عاملی

مقدمه

امروزه با گسترش تحقیقات در علوم پزشکی و اجتماعی، پژوهش، آموزش و مدیریت، بسیاری از پدیده‌ها از طریق ابزارهایی چون پرسش‌نامه‌ها، چک لیست‌ها و تست‌های آموزشی و یا ارزیابی‌های داوران مورد مطالعه قرار می‌گیرد [۱] تا بتوانند رفتار افراد یا جامعه مورد مطالعه را کمی‌سازی نمایند [۲]. از آنجا که بسیاری از تحقیقات بر اساس ابزارهای مذکور انجام می‌شود، پایایی (Reliability) و روایی (Validity) آنها امری اجتناب ناپذیر است و مقدار آن باید در حد قابل قبولی باشد [۲-۴]. بر اساس مطالعات انجام شده بیشتر مطالعات در گزارش روایی و پایایی با مشکل رو به رو هستند به نحوی که اطلاعات کافی در مورد ابعاد روان سنجی ابزارها و یا در به کار بردن تست‌های روان سنجی وجود ندارد. بر اساس

۱- استادیار اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، مرکز تحقیقات سیاستگذاری و ارتقاء سلامت، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی قم، قم، ایران
تلفن: ۳۷۸۴۲۲۲۸-۰۲۵، دورنگار: ۳۷۸۴۲۲۲۸-۰۲۵، پست الکترونیکی: beigi60@gmail.com
۲- کارشناس ارشد اپیدمیلوژی، مرکز تحقیقات سیاستگذاری و ارتقاء سلامت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی قم، قم، ایران
۳- دانشجوی دکتری آموزش بهداشت، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید بهشتی، تهران، ایران

[Downloaded from journal.rums.ac.ir on 2026-11-08]
[DOR: 20.1001.1.17353165.1393.13.12.4.7]

مطالعه‌ای مروری گرچه در برخی مطالعات روایی محتوی (Content Validity) بیان شده است اما این روایی با کمتر از ۵ متخصص انجام شده است. روایی ضابطه به ندرت گزارش شده است و یا خطاهایی در اندازه‌گیری ضابطه وجود دارد. همچنین، روایی سازه گزارش نمی‌شود. در خصوص پایایی بیشتر با استفاده از آلفای کرونباخ مقدار سازگاری درونی (Internal Consistency) گزارش می‌شود اما ثبات (Stability) آزمون‌ها کمتر مورد توجه قرار می‌گیرد [۵]. لذا بررسی ابعاد مختلف روانسنجی ابزارها و مقیاس‌ها یکی از ضروریات اندازه‌گیری در علوم مختلف و پژوهش است. لذا نوشتار حاضر با هدف آشنایی محققین و پژوهشگران دانشگاهی با مفاهیم روایی و پایایی و روش‌های مختلف اندازه‌گیری پایایی و همچنین، روایی به منظور استفاده صحیح و به جا از روش‌های مختلف تهیه شده است. در این مقاله ابتدا تعریفی از روایی و پایایی ارائه شده و به رابطه بین پایایی و روایی پرداخته شده است. سپس به پایایی و روش‌های اندازه‌گیری آن و انواع سنجش‌های پایایی شامل سنجش توافق درونی در پایایی با استفاده از آلفای کرونباخ ، پایایی دو نیمه آزمون و پایایی کادر ریچاردسن و همچنین، سنجش ثبات در پایایی با استفاده از آزمون - باز آزمون، فرم‌های یکسان، فرم‌های متعادل و پایایی بین مشاهده کنندگان اشاره شد. در نهایت به تعریف روایی، انواع روش‌های تعیین روایی شامل روایی صوری، روایی محتوی، روایی سازه و روایی ضابطه اشاره شد.

تعریف روایی و پایایی

روایی درجه‌ای از صحت نتایج مطالعه است و نشان می‌دهد که مطالعه تا چه اندازه‌ای چیزی را که قصد اندازه‌گیری آن را داشته است به درستی سنجیده است؟ [۱۱-۵، ۳-۱]. پایایی درجه‌ای از یکسان بودن نتایج در طول زمان معین و تحت شرایط مشابه و با روش کار مشابه می‌باشد که با قابلیت تکرار (Repeatability) و قابلیت تکثیر (Reproducibility) نتایج اندازه‌گیری می‌شود [۱۱-۵، ۳-۱] به عبارتی می‌توان گفت که پایایی شامل سه بخش متفاوت است. ثبات که به معنی کسب نتایج یکسان در افراد مشابه در صورت تکرار آزمون است؛ هم ارزی یا تعادل (Equivalence) هنگامی به دست می‌آید که چنانچه آزمون توسط محققین مختلفی به کار رود نتایج یکسانی را به دست آورد؛ و همگنی (Homogeneity) شاخصی است که تمام بخش‌های آزمون مورد استفاده دارای سازگاری درونی باشند. بنابراین برای تعیین همگنی، تعدادی از تعاریف مختلف عملیاتی از مفاهیم مشابه بر روی افراد مشابه با یک روش خاص در شیوه جمع‌آوری داده‌ها آزمون می‌شود و نتایج به دست آمده باید شدیداً به یکدیگر وابسته باشند [۱۳-۱۲].

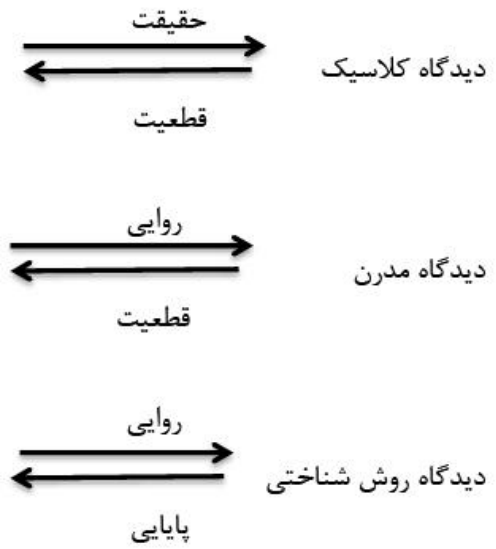
در ارزیابی‌های به کار رفته در علوم مختلف، در فعالیت‌های آموزشی و پژوهشی، تشخیص‌های پزشکی و غیره، بر اساس آن که محقق بخواهد هر کدام از اجزای پایایی شامل ثبات، هم ارزی و یا همگنی را اندازه‌گیری نماید می‌تواند از روش‌های مختلف به سنجش پایایی بپردازد. بدین منظور جهت بررسی پایایی ابزار، پرسش‌نامه و یا هر مقیاس سنجش دیگر می‌توان از روش آزمون - باز آزمون (Test-Retest) و همچنین، آزمون‌های مورد استفاده در تعیین پایایی از جمله ضریب آلفای کرونباخ، روش دو نیمه کردن، کادر ریچارسون و یا فرم‌های موازی بر روی نمونه‌ای از گروه هدف استفاده کند.

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

رابطه بین پایایی و روایی

روایی و پایایی سنجش‌ها، معیارهای ضروری در تعیین دقت و صحت سنجش می‌باشند. پایایی با خطای تصادفی در ارتباط است و روایی با خطای منظم و بنابراین هرچقدر که تعداد حجم نمونه بالا رود می‌تواند بر کاهش خطای تصادفی تأثیر گذارد و بنابراین پایایی ابزار افزایش می‌یابد و بنابراین بر دقت اندازه‌گیری مؤثر است. اما باید توجه نمود که افزایش روایی مستلزم استفاده از ابزارهای استاندارد و مؤثر خواهد بود. به بیان دیگر روایی نشان دهنده صحت اندازه‌گیری است [۱۵-۱۴]. ذکر این نکته لازم است که پایایی شرط لازم برای روایی است. به عبارتی چنانچه سنجه‌ای دارای پایایی نباشد نمی‌تواند مقدار درست و واقعی یک پدیده را نشان دهد. بنابراین در اعتبارسنجی ابزارها، سنجش روایی باید پس از تأیید پایایی صورت گیرد. پایایی با خطای تصادفی و روایی با خطای منظم اندازه‌گیری، مرتبط است بنابراین افزایش دقت با کاهش خطای تصادفی، پایایی را بالا می‌برد و افزایش صحت ناشی از کاهش خطای منظم بوده که روایی را بالا می‌برد [۱۶، ۱۴-۱۳]. شرط لازم برای روایی یک آزمون پایایی آن است ولی شرط کافی نیست و برای اینکه یک آزمون معتبر (روا) باشد باید پایا باشد [۱۷، ۱۴، ۱]. اما به طور کلی اعتبار و پایایی از هم جدا نبوده و با افزایش پایایی، روایی آن هم افزوده می‌شود [۱۹-۱۸] و البته پایین بودن پایایی نیز سبب کاهش روایی می‌گردد و نشان‌دهنده مجموعه‌ای از خطاهایی است که در اندازه‌گیری رخ می‌دهد [۱۹]. گرچه برخی محققین رابطه معکوسی بین پایایی و روایی را در ابزارها نشان داده‌اند، به عنوان مثال مصاحبه با سؤالات باز، پایایی کمی دارد حال آنکه روایی بالایی دارد اما یک مصاحبه استاندارد شده نتیجه عکس خواهد داشت [۲۰]. از منظر کلاسیک و روش شناختی نیز تفاوت قابل توجهی در مفاهیم بنیادی حقیقت (Truth) و قطعیت (Certainty) وجود دارد که در شکل ۱ مشخص شده است [۲۱-۲۰].



شکل ۱- نگاهی اجمالی به لغت شناسی مفاهیم پایه روایی و پایایی از دیدگاه‌های کلاسیک، مدرن و روش شناختی

از بعد روش شناختی روایی با حقیقت در ارتباط است و پایایی با قطعیت مرتبط می‌باشد. علاوه بر این رابطه معکوس بین روایی و پایایی در روش متدولوژیک و همچنین، رابطه معکوس بین حقیقت و قطعیت در دیدگاه کلاسیک نشان داده شده است. به این معنی که هر چه پایایی کمتر باشد عدم قطعیت افزایش می‌یابد [۲۱-۲۰]

پایایی و روش‌های اندازه‌گیری آن

ضریب پایایی عددی بین صفر تا یک می‌باشد که صفر نشانگر عدم وجود پایایی است و یک پایایی صد در صد را نشان می‌دهد [۲۲]. بنابراین پایایی شاخصی برای ارزشیابی آزمون‌ها و پرسش‌نامه‌ها اعم از آزمون‌های غربالگری و تشخیصی و پرسش‌نامه‌های تحقیقاتی می‌باشد و به دو دلیل از اهمیت خاصی بر خوردار است: اول آنکه

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

پایایی حاکی از وجود خطای تصادفی (Random Error) در اندازه‌گیری است که این خطا به دلایل فاکتورهای مربوط به آزمون شونده، فاکتورهای مرتبط با آزمون و شرایط برگزاری آن و همچنین، فاکتورهای مرتبط با امتیاز دهی ایجاد می‌شوند که اینگونه خطاها تصادفی بوده و اثر آنها بر روی امتیاز آزمون غیر قابل پیش‌بینی است. بر این اساس چنانچه حجم نمونه افزایش یابد و یا تعداد سؤالات یا آیتم‌های ابزار افزایش یابد پایایی بالا می‌رود. دوم، پایایی پیشگوکننده روایی است [۲۴-۲۳، ۹-۸، ۲].

انواع سنجش‌های پایایی

روش مورد استفاده برای سنجش پایایی یک ابزار (آزمون یا پرسش‌نامه) به دلیل عملکرد و طبیعت ابزار و چگونگی استفاده از آن در زمینه مورد استفاده متفاوت می‌باشد. گاهی هدف تعیین ضریب پایایی به معنای تعیین توافق درونی است که از روش‌های آلفای کرونباخ، کادر ریچارسون و دو نیمه کردن می‌توان استفاده کرد. اما چنانچه بخواهیم ثبات ابزار را مورد ارزیابی قرار دهیم می‌توان از روش آزمون بازآزمون و یا از روش فرم‌های موازی استفاده نمود [۲۵]. همچنین، گاهی هدف از تعیین ضریب پایایی تعیین توافق بین مشاهده‌گران یا مصاحبه کننده‌گان و یا ارزیابان می‌باشد.

سنجش توافق درونی در پایایی

۱- روش آلفای کرونباخ (Cronbach's Alpha) :

کرونباخ در سال ۱۹۵۱ در دانشگاه استنفورد روش آماری ضریب آلفا را برای حل مشکل تعیین پایایی آزمون‌های چند سؤالی ابداع کرد. این روش معمول‌ترین ضریب پایایی ثبات داخلی است که در بیشتر مطالعات از آن استفاده می‌شود [۳۵-۲۶] و معرف میزان تناسب گروهی از آیتم‌هایی است که یک‌سازه را می سنجند. مقدار آلفا باید حداقل برابر با ۰/۷ و یا بیشتر باشد تا یک سؤال بتواند در یک ابزار باقی بماند و بسیاری از محققان نقطه ۰/۸ را برای باقی ماندن یک سؤال در ابزار ضروری می‌دانند. پایایی برابر با ۰/۷، ثبات ۷۰ درصدی امتیازات یا نمرات به دست آمده از یک ابزار را نشان می‌دهد [۲۲، ۳۷-۳۶] و روشی برای محاسبه ثبات درونی بر اساس میانگین همبستگی درونی سؤالات می‌باشد و هنگامی که مقیاسی از نوع مقیاس لیکرت در ابزار اطلاعات به کار رود محاسبه مقدار ضریب آلفا جهت تعیین پایایی ثبات درونی امری ضروری است [۴۰-۳۸، ۳۶، ۲۲]. مقدار عددی آلفای کرونباخ برابر با متوسط مقدار ضریب پایایی است که از کل ترکیب‌های ممکن آیتم‌های ابزار، پس از آنکه به دو نیمه تقسیم شود به دست می‌آید [۴۲-۴۱، ۳۸، ۲۲]. روش آلفای کرونباخ یکی از روش‌های سنجش پایایی است که فقط به انجام یک بار آزمون نیاز دارد تا برآوردی از پایایی آزمون را فراهم کند [۲۳-۲۲]. درکل ضریب پایایی آلفای کرونباخ هنگامی مفید است که سؤالات یک ابزار به صورت صحیح – غلط امتیاز بندی نشده باشند و جهت سنجش ثبات درونی سؤالات در مقیاس لیکرت به کار می‌رود [۴۳، ۴۱، ۳۶، ۲۳]. در سال ۲۰۰۳ مقدار آلفای بیش از ۰/۹ را عالی، ۰/۹-۰/۸ را خوب، ۰/۸-۰/۷ را قابل قبول، ۰/۷-۰/۶ را قابل بحث، ۰/۶-۰/۵ را ضعیف و کمتر از ۰/۵را غیر قابل قبول معرفی کردند [۴۵-۴۴، ۳۷]. فرمول تعیین آلفای کرونباخ به قرار زیر است [۴۲، ۱۶]

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right),$$

در این فرمول K تعداد آیتم‌های ابزار S_i^2 واریانس آیتم i و S_t^2 واریانس کل ابزار یا تست (که از مجموع کل آیتم‌ها به دست آمده) می‌باشد [۴۲، ۲۵].

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

۲- پایایی دو نیمه آزمون (Split half reliability) :

در این روش یک آزمون به دو نیمه تقسیم شده و رابطه همبستگی بین دو نیمه آزمون سنجیده می‌شود. این آزمون معمول برای متغیرهای دو حالتی که کد صفر به پاسخ‌های غلط و کد یک به پاسخ‌های صحیح داده می‌شود به کار می‌رود. در این روش دو نیمه آزمون باید در محتوی و سختی سؤالات باهم مشابه باشند و از آنجا که تعداد سؤالات ابزار به دو قسمت تقسیم شده است می‌بایست روش ضریب همبستگی به کار رود که با استفاده از ضریب اسپیرمن و گاتمن محاسبه می‌شود [۴۶، ۴۱، ۲۵]. ضریب اسپیرمن براون فرمی از سنجش پایایی دو نیمه آزمون است که برای برآورد پایایی کل آزمون به کار می‌رود. همبستگی پیرسون بین دو نیمه آزمون، پایایی نیمه‌ای از ابزار یا مقیاس را برآورد می‌کند اما ضریب پیشگویی اسپیرمن، پایایی کل آزمون را بر اساس نیمه‌های آن برآورد می‌کند. مقدار این ضریب بیشتر از ضریب پایایی نیمه‌های آزمون (X_1 وX_2) است و معمولاً مساوی با دو برابر ضریب نیمه آزمون تقسیم بر مقدار یک به اضافه پایایی نیمه دیگر آزمون خواهد بود. ضریب پایایی اسپرمن براون با فرمول زیر محاسبه می‌شود [۲۵].

$$SB_2 r_{XX'} = \frac{2r_{X_1X_2}}{1 + r_{X_1X_2}}.$$

در بین سنجش‌های پایایی دو نیمه آزمون، ضریب پایایی اسپیرمن شدیداً تحت تأثیر ترتیب و نحوه قرار گرفتن سؤالات ابزار مورد استفاده در هر یک از دو نیمه آزمون می‌باشد. در این روش باید به شکل کاملاً دقیقی، تصادفی انجام شود و تخصیص تصادفی سؤالات بین دو نیمه آزمون باید تساوی واریانس بین دو فرم را تأمین نماید. اما همیشه این امر باید توسط محقق چک شود. زیرا هیچگاه این امر تضمین نمی‌شود و مقدار پایایی از یک اندازه‌گیری به اندازه‌گیری بعدی متفاوت خواهد بود [۴۷، ۲-۱]. ضریب گاتمن فرض می‌کند که پایایی نیمه اول و دوم (دو نیمه فرم) یکسان است و به تساوی واریانس‌های بین دو نیمه آزمون نیازی ندارد. بهترین نیمه هنگامی ایجاد می‌شود که هر نیمه سؤالاتی را که بیشترین همبستگی درونی دارند، شامل شود. به عبارتی زمانی که ابزاری به دو نیمه تقسیم شود به طوری که هر نیمه سؤالاتی را که شدیداً وابسته به همان گروه هستند و هیچگونه وابستگی بین دو گروه ندارند را شامل شود، ضریب دو نیمه گاتمن کمترین مقدار ممکن را خواهد گرفت و زمانی که سؤالات شدیداً وابسته به یکدیگر از هم جدا شده و در دو گروه متفاوت قرار گیرند ضریب دو نیمه گاتمن بشترین مقدار ممکن را خواهد گرفت [۴۱، ۱۶، ۱].

۳- پایایی کادر ریچاردسن (Kuder- Richardson reliability coefficient):

این روش شامل دو آزمون KR_{20} و KR_{21} می‌باشد. این روش‌ها، آزمون‌های همگنی (ثبات بین سؤالات) می‌باشند که نسبت پاسخ‌های صحیح به غلط را در هر سؤال ابزار یا آزمون مد نظر قرار می‌دهند و برای آزمون‌هایی که پاسخ آنها به شکل صحیح و غلط می‌باشد مفید است و در صورتی که که کد صفر به پاسخ‌های غلط و کد یک به پاسخ‌های صحیح داده می‌شود به کار رود [۴۸-۴۷]. از لحاظ ریاضی قابل اثبات است که ضریب پایایی کادر ریچاردسن برابر با میانگین ضرایب نیمه‌های آزمون به دست آمده از کل نیمه‌های ممکن یک آزمون می‌باشد در اینگونه آزمون‌ها (دو حالتی) مقدار KR_{20}، برابر با آلفای

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

کرونباخ می‌شود. اما روش KR_{21} هنگامی مفید است که مطمئن باشیم کل سؤالات پرسش‌نامه از لحاظ سختی با هم برابرند [۲۳، ۱۹، ۱۶، ۱]. در این روش آزمون یا ابزار هنگامی مورد پذیرش است که ضریب پایایی کادر ریچاردسن حداقل برابر با ۰/۶۴ باشد اما در آزمون‌های استاندارد این مقدار باید از ۰/۸۰ بالاتر باشد [۴۵، ۱۸].

$$KR\,20 = \left[\frac{n}{n-1}\right] \times \left[\frac{1-(\Sigma qp)}{Var_t}\right]$$

در این فرمول n تعداد آیتم‌ها، Var واریانس کل تست، Σqp مجموع ضرب درصدی از افرادی است که به آیتم پاسخ صحیح داده‌اند (p) در نسبتی از کل افرادی است که به آیتم پاسخ غلط داده‌اند (q) می‌باشد [۴۸-۴۷].

سنجش ثبات در پایایی

الف- پایایی آزمون - باز آزمون:

در این روش جهت سنجش پایایی، سؤالات آزمون در دو نوبت به یک گروه واحد، تحت شرایط مشابه داده می‌شود و نمرات حاصل با هم مقایسه می‌شوند. در نهایت ضریب همبستگی بین نمرات حاصل از اجرای دو بار آزمون برآورد شده تا چگونگی شباهت امتیازات مشخص گردد و به عنوان ضریب پایایی به کار می‌رود. معمول‌ترین روش برای یافتن همبستگی بین نمرات در این روش، محاسبه ضریب همبستگی اسپیرمن براون می‌باشد [۴۱، ۲۵]. روش آزمون- بازآزمون فرض می‌کند که متغیرها یا مفاهیم مورد اندازه‌گیری و همچنین، خصوصیات آزمون شوندگان در طول دوره تغییر نخواهند کرد [۴۹] و به دلیل برخی معایب آن از قبیل آشنایی آزمون شوندگان با محتوی سؤالات، مشکل در تعیین فاصله زمانی مناسب برای اجرای دو نوبت آزمون و عدم علاقه آزمون شوندگان به اجرای نوبت دوم توسط بسیاری از محققان بی‌ارزش دانسته شده است [۴۹، ۴۱]. محققین سه نوع اثر متفاوت شامل اثر حافظه (Memory Effect)، اثر خستگی (Fatigue Effect) و اثرات ناشی از تغییرات واقعی (Genuine Change Effect) در مقدار پایایی آزمون بازآزمون مؤثر خواهند بود [۲۳]. اثر حافظه که به دلیل یادآوری سؤالات در آزمون است سبب ایجاد نوعی خطای همبستگی (غیر تصادفی) شده و باعث بالا رفتن همبستگی نمرات آزمون و بازآزمون می‌گردد که پایایی را بالا می‌برد. اثر خستگی نوعی خطای تصادفی در بازآزمون شده و مقدار پایایی را کاهش می‌دهد. خطای ناشی از تغییرات واقعی نیز معمولا سبب کاهش پایایی می‌گردد [۲۳].

فاصله زمانی بین آزمون و باز آزمون یکی از چالش‌های اصلی در به کار بردن روش آزمون بازآزمون در تعیین پایایی سنجه‌ها است. فاصله زمانی مناسب برای انجام دو آزمون به ثبات متغیرهایی که عامل تحت بررسی را تحت تأثیر قرار می‌دهد بستگی دارد. به عنوان مثال چنانچه متغیر مورد نظر اندازه‌گیری فشار خون باشد زمان کوتاه خواهد بود اما جهت سنجش نگرش در مورد یک پدیده روانشناختی به فاصله زمانی طولانی تری نیاز است. اما فاصله زمانی متداول، چند هفته می‌باشد [۵۰، ۴۱، ۲۳، ۵]. بنابراین نکته مهم در این روش، فاصله زمانی بین دو آزمون نکته مهمی در استفاده از این آزمون است و اعتقاد بر آن است که فاصله زمانی بین دو آزمون باید تا حدی باشد که از طرفی فراموشی عبارات ابزار اتفاق بیفتد و از طرف دیگر تغییر در پدیده مورد اندازه گیری رخ ندهد برخی معتقدند که این فاصله را بین دو هفته تا یک ماه پیشنهاد کرده‌اند.

قابل قبول‌ترین آزمون جهت تعیین ثبات (آزمون پایایی) آزمون شاخص همبستگی درون خوشه‌ای

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

(Interclass Correlation Coefficient) است . چنانچه این شاخص بالاتر از ۰/۸ باشد میزان ثبات مطلوب است [۵۱].

ب- پایایی فرم‌های موازی (یکسان) و دقیقاً یکسان (Parallel & strictly parallel forms reliability) :

این دو مدل، روش‌هایی برای آزمون آماری تساوی میانگین‌ها و واریانس‌ها می‌باشند. روش دقیقاً یکسان فرض می‌کند که امتیازات سؤال درست میانگین و واریانس مشابهی دارند در حالی که مدل موازی فرض می‌کند که آنها واریانس مشابهی دارند اما لازم نیست که میانگین مشابهی داشته باشند. برآورد پایایی در روش موازی مساوی با آلفای کرونباخ می‌باشد و در مدل دقیقاً موازی بر اساس آلفای کرونباخ بوده ولی از تفاضل میانگین سؤالات به دست می‌آید [۵۲، ۲۵].

ج- پایایی فرم‌های متعادل (Equivalent forms reliability):

در این روش دو فرم جداگانه در مورد یک موضوع، به یک گروه مشابه داده می‌شود و هر فرم تعدادی از سؤالات آزمون را شامل می‌شود. به عبارتی دو آزمون معادل در مورد یک مفهوم یا متغیر به خصوص تهیه شده و در فاصله کوتاهی به یک گروه واحد داده می‌شود. ضریب همبستگی بین نمرات حاصل از این دو فرم برابر با پایایی فرم‌های متعادل می‌باشد. این روش به دلیل مشکل در طراحی و ساخت دو فرم یا نمونه سؤال همسان و مساوی در مورد یک مفهوم، عدم علاقه آزمون شوندگان برای پر کردن فرم دوم و یادگیری ناشی از فرم اول و تأثیر آن در تکمیل فرم دوم کاربرد زیادی ندارد [۴۷، ۲۵].

د- پایایی بین مشاهده کنندگان (Inter rater or observer reliability):

این ضریب ثبات و همبستگی نمرات بین مشاهده کنندگان یا داوران یا ارزیابان مختلف یک موضوع را می‌سنجد و ضریب توافق هم نامیده می‌شود که با ضریب کاپا نشان داده می‌شود [۵۳، ۱۹] و زمانی که یک ابزار یا پرسش‌نامه با سؤالات تشریحی، بر روی افراد مشابهی توسط دو یا چند مصاحبه گر یا پرسشگر به کار رود و بخواهیم به یک توافق کلی در مورد ابزار دست یابیم به کار می‌رود [۵۴، ۲۵، ۱۶]. ضریب کاپا بین صفر تا یک متغیر است و به صورت درصد بیان می‌شود و بر اساس مقالات موجود دسته بندی قدرت ضریب کاپا به قرار جدول ۱ است [۲۴].

جدول ۱- قدرت مقادیر مختلف ضریب کاپا در تعیین میزان توافق بین مشاهده کننده‌گان یا داوران

مقدار آماره کاپا	قدرت توافق
کمتر از صفر	ضعیف
۰-۰/۲	کم
۰/۲۱-۰/۴	پایین تر از متوسط
۰/۴۱-۰/۶۰	متوسط
۰/۶۱-۰/۸۰	خوب
۰/۸۱-۱	عالی

بر اساس مقادیر ارائه شده در جدول ۱ حداقل مقدار قابل قبول ضریب کاپا بیش از ۰/۶ می‌باشد و مقادیر بالاتر از ۰/۸ در توافق دو داور یا ارزیاب ایده‌آل می‌باشد.

همانطور که بیان شد ضریب پایایی می‌تواند با یک آزمون و یا بیش از یک آزمون محاسبه شود و همچنین، می‌تواند آزمون یا آزمون‌های مورد استفاده در یک زمان یا بیش از یک زمان به کار روند.

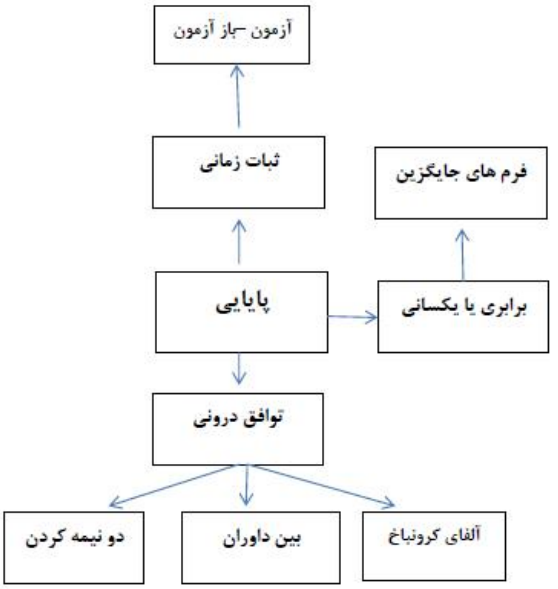
[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

جدول ۲- رابطه بین تعداد دفعات انجام آزمون در زمان‌های مختلف و تعداد فرم‌های یک آزمون در زمان خاص

دفعات لازم برای اجرای آزمون	تعداد سنجه‌های مورد نیاز جهت تعیین پایایی	
	یک آزمون	دو آزمون
یکبار	دو نیمه کردن کادر ریچارسون آلفای کرونباخ	فرم‌های موازی
دو بار	آزمون- بازآزمون	-----

در کل بر اساس تعداد دفعات انجام یک آزمون و تعداد فرم‌های آزمون مورد نیاز برای سنجش پایایی یک ابزار به دلیل عملکرد و طبیعت آن و چگونگی استفاده در طرح تحقیقاتی یا جامعه، باید یکی از روش‌های فوق را که در جدول ۲ و شکل شماره دو رابطه بین تعداد دفعات و تعداد فرم‌های یک آزمون را نشان می‌دهد، مورد استفاده قرار داد.



شکل ۲- انواع مختلف سنجش پایایی و ارتباط آنها با یکدیگر [۲]

بر اساس نمودار شماره دو نیز مشخص شده است که پایایی دو جزء متفاوت شامل ثبات و توافق درونی را شامل می‌شود که ثبات زمانی را با آزمون-بازآزمون می‌سنجند و توافق درونی را می‌توان با یکی از روش‌های دو نیمه کردن، توافق بین داوران با استفاده از ضریب کاپا و یا آلفای کرونباخ تعیین نمود.

تعریف روایی

سنجه‌های محقق ساخته و سنجه‌هایی که اعتبار آنها در فرهنگ‌ها و زبان‌های دیگر به تأیید رسیده باشد نیازمند اعتبار سنجی می‌باشند. در سنجه‌هایی با نسخ لاتین یا زبان بیگانه، ابتدا پرسش‌نامه توسط دو فرد متبحر (و یا بیشتر) از زبان اصلی به زبان فارسی ترجمه می‌شود. در مرحله بعدی طی مقایسه و بررسی دو متن (متن‌های) ترجمه شده از نظر کیفیت با یکدیگر مقایسه میشوند و مورد ارزیابی قرار می‌گیرند. در مواردی مشاوره با مترجمین و فردی انگلیسی زبان صورت می‌گیرد و در نهایت برای انتخاب مناسب‌ترین ترجمه برای عبارات، تلفیق و ترکیب ترجمه‌های اولیه صورت گرفته تا در مورد یک نسخه فارسی مشترک توافق گردد و بدین طریق پرسش‌نامه به یک ترجمه واحد تبدیل می‌شود. سپس نسخه نهایی ترجمه شده، جهت برگرداندن از زبان فارسی به زبان اصلی (ترجمه معکوس) در اختیار یک مترجم دیگر که متن اولیه انگلیسی را ندیده بود، قرار می‌گیرد. هدف از این کار اطمینان بیشتر از صحت ترجمه پرسش‌نامه است. پس از مقایسه نسخه ترجمه شده به انگلیسی با متن اصلی پرسش‌نامه و مشاوره با مترجم مربوطه و فرد انگلیسی زبان، یکسانی مفهومی (در بر داشتن محتوای مفهومی نسخه اصلی پرسش‌نامه) و کیفیت کلی ترجمه بررسی می‌گردد. سپس نسخه نهایی فارسی بر اساس ترجمه نهایی پرسش‌نامه معکوس و با نظر مترجمان و متخصصین تهیه می‌گردد. در نهایت جهت

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

تعیین روایی ابزار، روایی صوری (Face Validity)، روایی محتوی (Content Validity) و روایی سازه (Construct Validity) ابزار یا سنجه مورد استفاده باید بررسی و ارزیابی گردد. در پرسش‌نامه‌های محقق ساخته، پس از تهیه بانک سؤالات و تهیه پرسش‌نامه مورد نظر تیم پژوهش، نیز تعیین روایی صورت می‌گیرد. بدین صورت چه در سنجه‌های برگردان از زبان‌های دیگر و چه در سنجه‌های محقق ساخته، پس از تهیه نسخه نهایی فارسی پرسش‌نامه، در ابتدا روایی صوری بررسی می‌شود تا در صورت نیاز به تغییر پرسش‌نامه کل اعتبار پرسش‌نامه دستخوش تغییر نشود.

روایی صوری: روایی صوری قضاوت عینی در مورد ساختار ابزار است و از نگاه گروه هدف به این سؤالات پاسخ می‌دهد که: آیا ابزار طراحی شده به صورت ظاهری با هدف مطالعه مرتبط است؟ آیا افرادی که قرار است به ابزار پاسخ دهند، با عبارات و جمله‌بندی ابزار موافق هستند؟ آیا برداشت افراد غیر متخصص (گروه هدف) همان برداشتی است که مورد نظر محقق است؟ آیا اجزاء و کلیت ابزار قابل پذیرش پاسخ‌دهندگان می‌باشد [۲].

برای تعیین روایی صوری ابزار می‌توان از دو روش کیفی و کمی روایی صوری استفاده نمود. در روش کیفی و درآغاز فرآیند تعیین روایی صوری، با کمک متخصین امر در رشته‌های مرتبط با زمینه تحقیقاتی و پرسش‌نامه، از متخصصین آمار و اپیدمیولوژی و بر حسب موضوع از متخصصین روانشناسی یا علوم دیگر استفاده می‌شود. برای ارزیابی ضریب تمیز هر سؤال همبستگی آنها با نمره کل سنجیده می‌شود. بدین منظور پرسش‌نامه فارسی به دو فرد هیأت علمی متبحر در زبان انگلیسی و واژه‌های تخصصی داده می‌شود تا نسخه نهایی فارسی را از نظر مطلوب بودن عبارات به لحاظ وضوح (استفاده از واژه‌های ساده و قابل فهم)، کاربرد زبان مشترک (پرهیز از به کارگیری واژه‌های فنی و تخصصی) مورد بررسی قرار دهند. در صورت لزوم می‌توان تغییراتی را در سنجه مورد استفاده در جهت ساده تر کردن و قابلیت درک بیشتر انجام داد. همچنین، برای دریافت نظرات گروه هدف، مصاحبه‌ای با نمونه‌ای از آنها جهت یافتن دشواری در درک عبارات و کلمات، تناسب و ارتباط مطلوب آیتم‌ها، احتمال وجود ابهام و برداشت‌های نارسا از عبارات و یا وجود نارسایی در معانی کلمات انجام می‌شود و در صورت وجود مشکلات بایستی نظرات آنان به صورت تغییراتی جزئی در پرسش‌نامه اعمال شود.

روایی صوری سنجه‌ها به صورت کمی را با استفاده از روش تأثیر آیتم اندازه‌گیری می‌کنند. بدین منظور برای هریک از آیتم‌های سنجه، طیف لیکرتی ۵ قسمتی در نظر گرفته می‌شود: کاملاً مهم است (امتیاز ۵)، مهم است (امتیاز ۴)، به طور متوسطی مهم است (امتیاز ۳)، اندکی مهم است (امتیاز ۲) و اصلاً مهم نیست (امتیاز۱). سپس پرسش‌نامه جهت تعیین روایی صوری در اختیار گروه هدف قرار گرفته و پس از تکمیل پرسش‌نامه‌ها توسط گروه هدف، با استفاده از فرمول روش تأثیر آیتم، روایی صوری محاسبه می‌گردد.

$$Importance\ Score = \frac{\sum(f \times Importance)}{N}$$

f در فرمول تأثیر آیتم، فراوانی تکرار امتیاز توسط گروه هدف را نشان می‌دهد.

روایی محتوی

روایی محتوی بطور معمول به چنین سؤالاتی پاسخ می‌دهد که آیا ابزار طراحی شده همه جوانب مهم و اصلی مفهوم مورد اندازه‌گیری را در بردارد؟ آیا سازه‌های ابزار

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

همان چیزی را که باید بررسی می‌کند؟ آیا اجزاء و کلیت ابزار قابل پذیرش متخصصان ذیربط می‌باشد؟ جهت تعیین روایی محتوای می‌توان از دو روش کیفی و کمی روایی محتوا استفاده کرد [۵۵]. در روش کیفی سنجش اعتبار محتوا، می‌توان مصاحبه با متخصصین در پیامد مورد اندازه‌گیری و همچنین، مشاوره با متخصصین اپیدمیولوژی و آمار داشت. در ارزیابی کیفی روایی محتوا رعایت دستور زبان، استفاده از کلمات مناسب، اهمیت آیتم‌ها، قرارگیری آیتم‌ها در جای مناسب خود، زمان تکمیل ابزار طراحی شده مورد توجه قرار میگیرد. پس از جمع‌آوری ارزیابی متخصصین، با مشاوره با اعضای تیم تحقیق، تغییرات مورد نیاز در ابزار می‌بایستی مورد توجه قرار گیرد [۵۶-۵۵].

گرچه در بیشتر مطالعات تنها به ذکر این نکته بسنده می‌شود که روایی محتوی توسط متخصصین و صاحبنظران بررسی شد [۳۲-۳۰، ۲۸، ۲۶] اما باید توجه داشت که این شیوه از نظر متدولوژی خالی از اشکال نیست. روایی محتوی به صورت کمی نیز بر اساس نظرات متخصصین و با محاسبه دو شاخص نسبت روایی محتوا Content Validity Ratio (CVR) و شاخص روایی محتوا Content Validity Index (CVI) محاسبه می‌شود [۶۰-۵۷]. جهت اطمینان از اینکه مهمترین و صحیح‌ترین محتوا (ضرورت آیتم) انتخاب شده است از شاخص نسبت روایی محتوا و برای اطمینان از این که آیتم‌های ابزار به بهترین نحو جهت اندازه‌گیری محتوا طراحی شده‌اند از شاخص روایی محتوا استفاده می‌گردد. جهت تعیین نسبت روایی محتوی به صورت کمی از تعدادی از متخصصین و صاحبنظران استفاده می‌شود تا در مورد هر کدام از آیتم‌های ابزار یا سنجه مورد استفاده به سه طیف آیتم شامل ضروری است، مفید اما ضروری نیست و ضرورتی ندارد، پاسخ دهند. بهتر آن است که از تعداد ارزیابان بیشتری (بالاتر از ۱۰ نفر) استفاده شود. CVR بر اساس نظر داوران به قرار ذیل محاسبه می‌شود:

$$CVR = \frac{(ne - \frac{N}{2})}{\frac{N}{2}}$$

در فرمول نسبت روایی محتوی، ne تعداد ارزیابانی است که گویه مورد نظر را ضروری یا سودمند می‌دانند و N تعداد کل ارزیابان یا داورانی است که یک گویه را بررسی کرده‌اند [۵۹، ۵۷]. حداقل مقادیر CVR در تست‌های یک طرفه مورد نظر Lawshe در جدول ۳ آمده است [۶۱-۵۹].

جدول ۳- حداقل مقادیر CVR در تست‌های یک طرفه مورد نظر Lawshe جهت مناسب بودن روایی محتوی

تعداد ارزیابان	حداقل مقدار نسبت روایی محتوی
۵	۰/۹۹
۶	۰/۹۹
۷	۰/۹۹
۸	۰/۷۵
۹	۰/۷۸
۱۰	۰/۶۲
۱۵	۰/۴۹
۲۰	۰/۴۲
۲۵	۰/۳۷
۳۰	۰/۳۳
۳۵	۰/۳۱
۴۰	۰/۲۹

پس از تعیین و محاسبه CVR، می‌توان شاخص CVI را محاسبه نمود. برای محاسبه این شاخص، ارزیابان می‌بایست به هر آیتم ابزار مورد استفاده، در خصوص سه معیار مربوط یا اختصاصی بودن، سادگی و روان بودن و وضوح یا شفاف بودن، بر اساس طیف لیکرتی ۴ قسمتی

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

اظهار نظر می‌نمایند. مثلاً جهت معیار مربوط بودن گزینه‌های، مربوط نیست=۱، نسبتاً مربوط است =۲، و مربوط است =۳ و کاملاً مربوط است =۴، به کار گرفته می‌شود و سپس با استفاده از فرمول CVI شاخص روایی محتوا محاسبه می‌گردد [۶۴-۶۲، ۵۹، ۵۷].

$$Cv1 = \frac{\text{نسبت تعداد ارزیابی که به آیتم نمره ۳ و ۴ داده‌اند}}{\text{تعداد کل ارزیابان}}$$

روایی سازه: مفهوم مرکبی است که مستلزم بررسی چند مرحله است و با روایی ضابطه که شامل روایی همزمان، روایی پیشگویی کننده ، روایی تمایز و روایی همگرایی است سنجیده می‌شود [۲].

روایی ضابطه (Criterion-related Validity): درجه‌ای از توافق بین یک تست با یک یا تعداد بیشتری تست معتبر دیگر در همان زمینه (که بهتر است دقیق و ایده آل باشد) است که معمولاً با استفاده از ضریب همبستگی بین آنها سنجیده می‌شود [۲].

الف- روایی همزمان (Concurrent Validity): هنگامی برآورد می‌شود که معیار (استاندارد) در حال حاضر وجود دارد. به عنوان مثال می‌توان در یک بررسی از کارمندان میزان حقوق دریافتی را پرسید و به عنوان یک ضابطه از مدارک حسابداری به عنوان ضابطه استفاده نمود و همبستگی مقدار حقوق گزارش شده با مقدار واقعی آن بر اساس مدارک موجود را محاسبه کرد.

ب- روایی پیشگویی کننده (Predictive Validity): هنگامی کاربرد دارد که معیار در حال حاضر وجود ندارد و در زمان آینده اتفاق می‌افتد. هدف آن است تا بدانیم یک مقیاس یا ابزار تا چه اندازه می‌تواند پیامدی را در آینده پیش‌بینی کند، به عنوان مثال بر اساس نمره میان ترم دانشجویان قبولی و یا رد شدن آنها را در پایان ترم پیش‌بینی کنیم [۲].

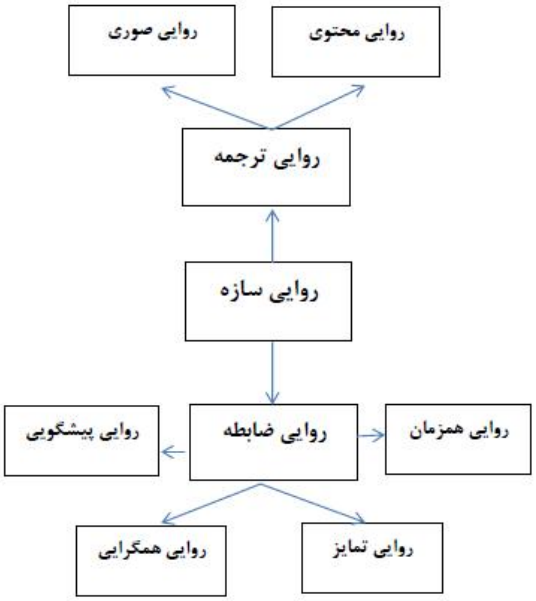
ج- روایی همگرایی (Convergent Validity) که عبارتست از همبستگی نسبتاً قوی میان سؤال مربوط به یک حیطه با همان حیطه که بنابر توصیه سازمان اروپایی تحقیقات و درمان سرطان ضریب همبستگی پیرسون باید بزرگتر از ۰/۴ باشد [۶۳، ۲]. اصل در جهت گروه‌بندی متغیرها نیز در یک حیطه خاص بر این منوال است که چنانچه ضریب همبستگی جدول ماتریس همبستگی کمتر از ۰/۳± باشد حداقل بوده، ۰/۴± بیانگر مهم بودن متغیر و ۰/۵± بیانگر اهمیت اساسی متغیر در حیطه مورد نظر است. اما چنانچه ضریب همبستگی کمتر از ۰/۳± محاسبه شود محقق باید بررسی کند که آیا کاربرد تحلیل عاملی (در ادامه مقاله به آن پرداخته می‌شود) برای ابزار مناسب بوده یا خیر؟ [۵۸].

د- روایی تمایز (Discriminant Validity) که بر همبستگی ضعیف بین سؤالات مربوط به یک حیطه با دیگر حیطه‌ها دلالت دارد. به عبارت دیگر ضریب همبستگی یک سؤال با حیطه‌های دیگر باید کمتر از ضریب همبستگی آن سؤال با حیطه‌های مربوط به خودش باشد. یک ابزار موفق علاوه بر آنکه باید با ابزارهای دیگر که سازه مشابه را اندازه‌گیری می‌کند همگرا باشد باید با ابزارهای یا مقیاس‌های مفاهیم دیگری که یک نظریه را با مفهوم کانونی پیوند می‌زند نیز همبستگی بالایی داشته باشد. علاوه بر آن هرچند اعتبار همگرا لازم است اما کافی نیست و اعتبار تفکیکی در حد اعتبار همگرا لازم است زیرا به بررسی حساسیت ابزارها در سنجش پدیده ای غیر مربوط می پردازد و میزان ابهام یا گسترده بودن مفهوم را

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

نشان می‌دهند [۶۳، ۵۸، ۲]. شکل ۳ نشان دهنده انواع زیرشاخه‌های روایی است.



شکل ۳- روش‌های مختلف سنجش روایی سازه و ابعاد مختلف روایی ضابطه و ترجمه

بنابراین روایی سازه منتج از روایی ترجمه و روایی ضابطه می‌باشد که روایی ترجمه با تعیین روایی صوری و روایی محتوی به دست می‌آید و روایی ضابطه شامل چهار بعد روایی همزمان، روایی همگرایی، روایی تمایز و روایی پیشگویی می‌باشد [۲].

روایی ساختار (Construct Validity) آیتم‌های یک ابزار می‌تواند بر اساس تحلیل عاملی تعیین شود. تحلیل عاملی به دو شکل کلی انجام می‌گیرد که شامل تحلیل عاملی اکتشافی و تأییدی است. تحلیل عاملی اکتشافی در بیشتر نرم‌افزارهای آماری در دسترس از جمله SPSS قابل اجرا است حال این که تحلیل عاملی تأییدی با SPSS قابل انجام نبوده و باید از سایر نرم‌افزارها برای اجرای آن استفاده نمود. تحلیل عاملی تأییدی هنگامی کاربرد دارد که ابزار یا پرسش‌نامه مورد استفاده از روایی سازه مناسب برخوردار باشد و به منظور تأیید کارهای قبلی به کار می‌رود بنابراین امکان استخراج فاکتورهای جدید در این نوع تحلیل عاملی وجود ندارد. تحلیل عاملی اکتشافی معمولا با روش مؤلفه‌های اصلی (Principle Component Analysis (PCA) انجام می‌گیرد و ارتباط درونی میان متغیرها را مورد بررسی قرار می دهد تا طبقاتی از متغیرها که دارای بیشترین ارتباط با یکدیگر هستند را کشف نماید [۶۵، ۵۸، ۱۱]. به این صورت آیتم‌هایی از ابزار که در هر فاکتور بیشترین همبستگی را با یکدیگر نشان می‌دهند می‌توانند آیتم‌های مورد استفاده در تبین هر فاکتور یا مؤلفه ابزار قرار گیرند. جهت بررسی میزان تطابق و نامگذاری عوامل استخراج شده از چرخش واریماکس و یا پروماکس استفاده می‌گردد [۶۸-۶۶، ۵۹-۵۸].

نکته مورد توجه در تحلیل عاملی پرداختن به نمونه گیری کیسر مایر الکین (KMO) و آزمون کرویت بارتلت (BT) و همچنین مقدار عددی نقطه عطف یا ارزش ویژه Egenvalue است که معمولاً برابر با ۱ در نظر گرفته می‌شود و به عنوان حداقل بار عاملی مورد نیاز برای حفظ هر عبارت در عوامل استخراج شده از تحلیل عاملی در نظر گرفته می‌شود. قابل ذکر است که در تحلیل عاملی، مقدار Determinant می‌بایست مخالف صفر و مقدار کای اسکوئر آزمون بارتلت عددی بزرگ و معنی‌دار از نظر آماری باشد و همچنین معیار KMO به عدد یک نزدیک باشد تا نشان‌دهنده کفایت نمونه‌گیری و حجم نمونه کافی برای اجرای تحلیل عاملی مدنظر قرار گرفته شود. در خصوص حجم نمونه لازم جهت انجام تحلیل عاملی تناقضات متفاوتی وجود دارد. Tabachnick حداقل ۳۰۰ نمونه را برای انجام تحلیل عاملی ضروری می‌داند اما Hair و همکاران [۶۹] بیان می‌کنند که حجم نمونه باید بیش از

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

۱۰۰ باشد اما برخی منابع دیگر از جمله Comrey و Lee [۷۰] طبقه‌بندی انجام داده‌اند که ۱۰۰ را ضعیف، ۲۰۰ را نسبتاً خوب، ۳۰۰ را خوب، ۵۰۰ را خیلی خوب و ۱۰۰۰ و بیشتر را عالی معرفی کرده‌اند [۷۰، ۶۸-۶۶، ۵۸].

در آخر ذکر این نکته ضروری است که روایی و پایایی از شاخص‌های لازم جهت اندازه‌گیری و دستیابی به مقادیر واقعی پیامدها می‌باشد. آگاهی و فهم کافی از آنها یکی از اصول متدولوژی در طراحی مطالعات و سنجش‌های مختلف است که متاسفانه در بسیاری موارد نادیده گرفته می‌شود. بر اساس موارد ذکر شده در تعیین پایایی می‌توان بیان کرد که افزایش تعداد سؤالات مشابه با سؤالات قبلی برای طولانی‌تر کردن آزمون، متجانس و همگون کردن سؤالات، افزودن سؤالات با ضریب دشواری متوسط، افزودن سؤالات با ضریب تمیز بالا، به کارگیری آزمون بر روی گروهی از افراد که از نظر توانایی پاسخگویی به آزمون تجانس کمی‌دارند و همچنین، به کارگیری آزمون سرعت، یعنی اجرای آزمون در یک زمان محدود از عوامل مؤثر در افزایش پایایی ابزارها یا سنجه‌ها می‌باشند [۴۸-۴۷، ۲۳، ۱]. سنجش روایی نیز نیازمند آگاهی از علم آمار و روش‌های آماری است تا در دست‌یابی به مقدار واقعی پیامدها مفید واقع شود. بنابراین برنامه‌ریزی در جهت طراحی و کاهش خطاهای غیر تصادفی در اندازه‌گیری می‌تواند به پایایی و روایی ابزار اضافه کرده و قدرت نتایج مطالعه را افزایش دهد. کم گزارش دهی و عدم گزارش روایی ممکن است به دلایلی از جمله کم بودن حجم نمونه، طراحی ضعیف مطالعه و یا کاهش منابع باشد [۵]. همچنین، از آنجا که هدف نوشتار حاضر آشنایی محققین و دانشگاهیان با مفاهیم روایی و پایایی و شیوه‌های مختلف اندازه‌گیری آن است سعی شد تا به ساده‌ترین وجه ممکن مطالب مورد نیاز با یک روال منظم و منطقی بیان گردد. لذا از ذکر فرمول‌های پیچیده واریانس ضرایب پایایی و یا مفاهیم آماری مربوط به تحلیل عاملی صرف نظر شد.

نتیجه‌گیری

پایایی درجه‌ای از ثبات نتایج طی زمان و قابلیت تکرار آنها می‌باشد که سنجش پایایی علاوه بر روش معمول آلفای کرونباخ با روش‌های دیگر نیز قابل اندازه‌گیری است. در حالی که روایی درجه‌ای از صحت نتایج می‌باشد و ابعاد گوناگونی دارد. لذا توجه محققین و پژوهشگران و دست‌اندرکاران نظام سلامت به ارائه نتایج و آمار دقیق و صحیح می‌تواند در نتیجه استفاده از سنجش‌های پایا و قابل اعتماد به دست آید. سنجش‌هایی که علاوه بر کم کردن خطاهای تصادفی از خطای منظم نیز بتواند دوری کند.

تشکر و قدردانی

بدینوسیله از همکار محترم جناب آقای دکتر شهرام ارسنگ جنگ و مهندس محسن مختاری تشکر و قدردانی می‌نماییم.

References

[1] Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American j of Med* 2006; 119(2): 166, 7-16.

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

[2] Drost EA. Validity and Reliability in Social Science Research. *Education Research & Perspectives* 2011; 38(1):105-123

[3] Golafshani N. Understanding reliability and validity in qualitative research. *The qualitative report* 2003; 8(4): 597-607.

[4] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use: Oxford university press; 2014;

[5] DeVon HA, Block ME, Moyle Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *J Nursing Scholarship* 2007; 39(2): 155-64.

[6] Amiri P, Ardekani EM, Jalali-Farahani S, Hosseinpanah F, Varni JW, Ghofranipour F, et al. Reliability and validity of the Iranian version of the Pediatric Quality of Life Inventory™ 4.0 Generic Core Scales in adolescents. *Quality of Life Research* 2010; 19(10): 1501-8.

[7] Cuthbert SC, Goodheart GJ. On the reliability and validity of manual muscle testing: a literature review. *Chiropractic & Manual Therapies* 2007; 15(1): 4 doi:10.1186/1746-1340-15-4.

[8] Healy M, Perry C. Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualit Mark Res: An int J* 2000; 3(3): 118-26.

[9] Litwin MS. How to measure survey reliability and validity. *Sage*; 1995;

[10] Pearson Executive Office . PCAT Reliability and Validity 2015-2016 . San Antonio, Product Number 65685PCAT: 2012.

[11] Pintrich PR, Smith DA, García T, McKeachie WJ. Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement* 1993; 53(3): 801-13.

[12] Quinn TJ. Validity and Reliability of the Outdoor Education Practicum Questionnaire [microform] / Thomas James Quinn. [Washington, D.C.]: Distributed by ERIC Clearinghouse; 1998.

[13] Holbert RL, Lee J, Esralew S, Walther WO, Hmielowski JD, Landreville KD. Affinity for political humor: An assessment of internal factor structure, reliability, and validity. Humor. 2013;26(4):551-72

[14] Lang W, Wilkerson J. Accuracy vs. Validity, Consistency vs. Reliability, and Fairness vs. Absence of Bias: A Call for Quality. Annual Meeting of the American Association of Colleges of Teacher Education (AACTE); New Orleans, LA.2008.

[15] Mohammadbeigi A, Tabatabaee S, Mohammadsalehi N, editors. Measurment of Tests's Reliability 8th National Conference of Medical Education, 2006; Kerman.

[16] . Wells CS, Wollack JA. An instructor's guide to understanding test reliability. Testing & Evaluation Services publication, University of Wisconsin Retrieved January. 2003;4:2006.

[17] Carmines EG, Zeller RA. Reliability and validity assessment. *Sage* 1979.

[18] Seif AA. Evaluation and assessment of educational progeression. Agah Publication, Tehran, 1992

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

[19] Downing SM. Reliability: on the reproducibility of assessment data. *Medical Education* 2004; 38(9): 1006-12.

[20] Lafaille R, Wildeboer H. Validity and Reliability of Observation and Data Collection in Biographical Research: International Institute for Advanced Health Studies v.z.w. *Antwerp* 1995.

[21] Cuttance P, Ecob R. Structural modeling by example: Applications in educational, sociological, and behavioral research: Cambridge University Press, New york. 2009

[22] Gliem JA, Gliem RR, editors. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales2003: *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*.2005..

[23] Sechrest L. Reliability and validity *Research Methods in Clinical Psychology* 1984: 24-54.

[24] Gwet KL. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. *Advanced Analytics LLC* 2014

[25] Webb NM, Shavelson RJ, Haertel EH. Reliability coefficients and generalizability theory. *Handbook of Statistics* 2006; 26: 81-124.

[26] Ghamari F, Mohammadbeigi A, Mohammadsalehi N, Hashiani A. Internet addiction and modeling its risk factors in medical students, Iran. *Indian J Psychol Med* 2011 ;33:1562-68..

[27] Mohammadbeigi A, Mohammadsalehi N. Prevalence of Internet Addiction and Related Risk Factors in Students. *J Guilan Univ Med Sci* 2011; 20(78): 41-8.

[28] Abolfazl M, Hamidreza TS, Narges M, Maryam Y. Gestational diabetes and its association with unpleasant outcomes of pregnancy. *Pak J Med Sci* 2008; 24: 566-70.

[29] Ahmadi A, Ershad M, Givzadeh H, Mohammad-Beigi A. General physicians' knowledge about nutrition in Shiraz, Iran. *Pakistan J of Biol Sci* 2009; 12(13): 981-5.

[30] Bayati A, Beigi M, Salehi M. Depression prevalence and related factors in Iranian students. *Pak J Biol Sci* 2009; 12(20): 1371-5.

[31] Beigi A. Musculoskeletal symptoms among workers of metal structure manufacturing industry in Shiraz, 2005. *Iran J Epidemio* 2009; 5(3): 12-7.

[32] Beygi A, Salehi N, Bayati A. Prevalence of unintended pregnancy and its related factors in Arak 2007. *Koomesh* 2009; 10(3): 201-6.

[33] Hassanzadeh J, Mohammadbeigi A, Eshrati B, Rezaianzadeh A, Rajaeefard A. Determinants of Inequity in Health Care Services Utilization in Markazi Province of Iran. *Iran Red Crescent Med J* 2013; 15(5): 363.

[34] Mohammad BA, Tabatabaei S, Mohammad SN, Yazdani M. Factors influencing cesarean delivery method in Shiraz hospitals. *Iran J Nursing (IJN)* 2009; 21(56): 37-45.

[35] Tabatabaee HR, Mohammad-Beigi A, Yazdani M, Zeighami B, Mohammad-Salehi N. Gestational diabetes risk factors modeling in pregnant women. *Int J of Diab Develop Coun* 2007; 27(1): 11-3.

[36] Helms JE, Henze KT, Sass TL, Mifsud VA. Treating Cronbach's alpha reliability coefficients as data in

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

counseling research. *The Counseling Psychologist* 2006; 34(5): 630-60.

[37] DeVellis RF. Scale development: Theory and applications. Sage Publications 2011;

[38] Cronbach L, Meehl P. Construct validity in psychological tests'. *Personality: Critical Concepts* 1998; 52: 135.

[39] Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological bulletin* 1955; 52(4): 281.

[40] Ghasemi V. Cronbach Coefficient Alpha and some of Its Specialties With Em phasis on Applications in Social Research. *Humanities* 2005; 19(2): 155-74.

[41] Guttman L. A basis for analyzing test-retest reliability. *Psychometrika* 1945; 10(4): 255-82.

[42] Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement* 2004; 64(3): 391-418.

[43] Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Edu* 2011; 2: 53.

[44] Sijtsma, K., On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 2009; 74(1): 107-20.

[45] George D, Mallery M .Using SPSS for Windows step by step: a simple guide and reference. Boston, MA: *Allyn & Bacon* 2003;

[46] North Carolina State University. Introduction to Reliability. 2012. Available at http://www-4.ncsu.edu/~jlnietfe/ EDP560_Notes_files/Reliabilty.pdf. Access date; 04/04/2014.

[47] Introduction to Reliability. http://www4.ncsu.-edu/~jlnietfe/EDP560_Notes_files/Reliabilty.pdf. Access date; 2014.04.04.

[48] Miller IW, Epstein NB, Bishop DS, Keitner GI. The McMaster Family Assessment Device: Reliability and Validity. *J Marital and Family Therapy* 1985; 11(4): 345-56.

[49] Trochim WM. An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning* 1989; 12(1): 1-16.

[50] Knapp TR, Brown JK. Ten measurement commandments that often should be broken. *Research in Nursing & Health*. 1995; 18(5): 465-9.

[51] Najafi F, Kheyri B. Investigating the Impact of Country of Origin on Customer Behavior: Investigation of the Moderating Roles of Product Involvement and Product Familiarity on Product Evaluation and Customer Behavioral Intentions. *J Marketing Manangment* 2013; 17: 37-60.

[52] Kristof W. The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika* 1963; 28(3):221-38.

[53] Fleiss JL, Levin B, Paik MC. The measurement of interrater agreement. Statistical methods for rates and proportions. 1981; 2: 212-36.

[54] Kundel H, Polansky M. Measurement of observer agreement. *Radiology* 2003; 228: 303-8.

[55] Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health* 2006; 29(5): 489-97.

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

[56] Yaghmale F. Content validity and its estimation. *J Med Edu* 2009; 3(1): 25-7.

[57] Heravi-Karimooi M, Anoosheh M, Foroughan M, Sheykhi MT, Hajizadeh E. Designing and determining psychometric properties of the Domestic Elder Abuse Questionnaire. *Iran J Ageing* 2010; 5(15): 7-21.

[58] Williams B, Brown T, Onsman A. Exploratory factor analysis: A five-step guide for novices. *J Emergency Primary Health Care* 2012; 8(3): 1.

[59] Lawshe CH. A Qualitative Approach to Content Validity. *Personnel Psychology* 1975; 28(4): 563-75.

[60] Newman I, Lim J, Pineda F. Content Validity Using a Mixed Methods Approach: Its Application and Development Through the Use of a Table of Specifications Methodology. *J Mixed Methods Research* 2013.

[61] Hassanzadeh Rangi N, Allahyari T, Khosravi Y, Zaeri F, Saremi M. Development of an Occupational Cognitive Failure Questionnaire (OCFQ): Evaluation validity and reliability. *Iran Occupational Health* 2012; 9(1): 29-40.

[62] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health* 2007; 30(4): 459-67.

[63] Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research* 2003; 27(2): 94-104.

[64] Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. *West J Nurs Res* 2003; 25(5): 508-18.

[65] Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods* 2004; 7(2): 191-205.

[66] Tabachnick BG, Fidell L. Using Multivariate Statistics: New York: Harper & Row.; 2012.

[67] Tabachnick BG, Fidell LS, Osterlind SJ. Using multivariate statistics New York: Harper & Row 2001.

[68] Zare-Chahoki M. Multivariate Analysis Method In SPSS Software 2010. Tehran: Tehran University; Available at URL: http://utcan.ut.ac.ir/-member/syllabus/ mazare/multivariate.pdf

[69] Hair JF. Multivariate data analysis. Upper Saddle River, NJ Pearson Prentice Hall, New York; 2006.

[70] Comrey, Andrew L., and Howard B. Lee. A first course in factor analysis. Academic Press, New York, 2013.

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]

Validity and Reliability of the Instruments and Types of MeasurmentS in Health Applied Researches

A. Mohammadbeigi[1] , **N. Mohammadsalehi**[2], **M. Aligol**[3]

Received: 28/08/2014 Sent for Revision: 26/01/2015 Received Revised Manuscript: 01/02/2015 Accepted: 25/02/2015

Background and Objective: Nowadays, by extension of research in medical and social sciences, education and managing most of outcomes measurers by cheklists,questionnaires, scales and educational tests or raters evaluations to quantify the human behaviors. Reliability and validity are the most important indexes in measurement to obtain true values. Enough knowledge and understanding the reliability and validity are important principle of methodology in study designing. Unfortunately, psychometric aspects of measures have not been applied in some situations and studies. The objective of current document is to show the concepts of reliability, validity, psychometric aspects and their types of measurements for researchers .

Key words: Validity, Reliability, Psychometric , Cranach's Alpha, Kappa Coefficient, Factor Analysis

Funding: This study was funded by Qom University of Medical Sciences.

Ethical approval: The Ethics Committee of Qom University of Medical Sciences approved the study.

Conflict of interest: none declared.

How to cite this article. Mohammadbeigi A, Mohammadsalehi N, Aligol M. Validity and Reliability of the Instruments and Types of Measurment in Health Applied Researches. *J RafsanjanUniv Med Sci* 2015; 13(10): 1153-70. [Farsi]

1- Assistant Prof., Dept. of Epidemiology and Biostatistics, Health Policy and Promotion Research Center, Qom University of Medical Sciences, Qom, Iran
(Corresponding Author), Tel: (025)37842228, Fax: (025) 37842228, E-mail: Beigi60@gmail.com
2- MSC of Epidemiology, Health Policy and Promotion Research Center, Qom University of Medical Sciences, Qom, Iran
3- PhD Student of Health Education and Promotion, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[Downloaded from journal.rums.ac.ir on 2026-11-08]

[DOR: 20.1001.1.17353165.1393.13.12.4.7]