



فرایند ارزیابی روایی و پایایی ابزارهای سنجش و اندازه‌گیری در تحقیقات آموزشی، رفتاری و علوم سلامت (با تأکید بر روایی محتوا)

محمد مسعود وکیلی^{۱*}، نسرین جهانگیری^۲

Vakili@zums.ac.ir

^۱نویسنده‌ی مسوول: زنجان، دانشگاه علوم پزشکی زنجان، دانشکده‌ی بهداشت و پیراپزشکی

چاپ: ۱۳۹۶/۱۲/۲۰

پذیرش: ۱۳۹۶/۱۰/۲۰

دریافت: ۱۳۹۶/۰۷/۲۱

چکیده

بخش مهمی از تحقیقات علوم اجتماعی، آموزشی و پزشکی، به طراحی و ارزیابی ابزارهای سنجش و اندازه‌گیری صفات و متغیرهای کیفی و انتزاعی، مرتبط است. دو جزء اساسی و مهم این‌گونه ابزارهای محقق ساخته، عبارتند از روایی و پایایی. در استفاده از این نوع ابزارهای تحقیق، کیفیت ارزیابی و تأمین این دو ویژگی همواره به عنوان یک نگرانی مهم مطرح هستند. پژوهشگران قبل از انتشار نتایج مطالعات خود، ملزم به گزارش چگونگی ارزیابی کیفیت روایی و پایایی ابزارهای اندازه‌گیری هستند. اشاره دقیق به چنین خصوصیتی، قابلیت اظهار نظر درباره اعتبار و میزان صحت نتایج تحقیق و مقایسه با نتایج سایر مطالعات را فراهم می‌نماید. در صورتی که اعتبار و اعتماد ابزار تحقیق، مورد تردید قرار گیرد، ممکن است تلاش‌های محقق و منابع تحقیق، به هدر رود. هدف از نگارش این مطالعه، بیان اهمیت و جایگاه این دو ویژگی در تحقیقات مختلف از جمله علوم پزشکی و مراحل انجام روایی محتوا به دو روش کمی و کیفی، و دو روش رایج ارزیابی پایایی ابزارهای اندازه‌گیری، می‌باشد. **واژگان کلیدی:** روانسنجی، روایی محتوا، روایی صوری، پایایی، پرسشنامه.

این مقاله بدین صورت ارجاع داده شود:

Vakili M M, Jahangiri N. Content Validity and Reliability of the Measurement Tools in Educational, Behavioral, and Health Sciences Research . J Med Educ Dev. 2018; 10 (28) :106-118

۱- گروه آموزش بهداشت و ارتقای سلامت، دانشکده بهداشت و پیراپزشکی، دانشگاه علوم پزشکی زنجان، زنجان، ایران.

۲- مرکز مطالعات و توسعه آموزش، واحد پژوهش در آموزش، دانشگاه علوم پزشکی زنجان، زنجان، ایران.

مقدمه

به طور مستمر بر اهمیت ارزیابی و گزارش دقیق روایی و پایایی ابزارها و مقیاس‌های سنجش و اندازه‌گیری تأکید می‌کنند، اما نسبت بسیار زیادی از مطالعات منتشر شده در حیطه‌های مرتبط با علوم رفتاری و آموزش بهداشت، گزارش‌های مرتبط با روایی (مابین ۴۰ تا ۹۳ درصد) و پایایی (مابین ۳۵ تا ۸۰ درصد) ابزار تحقیق را گزارش نمی‌کنند که چنین آماری فاجعه‌بار است (۱). البته چنین نقصی صرفاً محدود به تحقیقات آموزش‌بهداشت و علوم اجتماعی و رفتاری نیست و با توجه به وسعت کاربرد ابزارهای سنجش و اندازه‌گیری در علوم سلامت، در بخش قابل‌توجهی از مطالعات انجام شده در سایر علوم نیز، ممکن است مشاهده شود. بدون ارزیابی چنین خواص مهمی از ابزار اندازه‌گیری، این امکان وجود دارد که پژوهشگران به سوی اخذ نتیجه‌گیری‌ها و توصیه‌های پژوهشی نادرست و نامعتبر سوق داده شوند. به عبارت دیگر، بدون حصول اطمینان از اینکه آیا ابزار مورد استفاده در یک مطالعه، توانسته است نمرات دقیق و پایدار ارائه دهد یا خیر، پژوهشگران نخواهند توانست یقین پیدا کنند که صفت، رفتار و یا متغیرهای مورد مطالعه، به درستی گزارش شده باشند. به منظور اینکه متخصصین و پژوهشگران، بتوانند سودمندی و قابلیت بکارگیری نتایج مطالعات خود را افزایش دهند، ضروری است توجه به ویژگی‌های روانسنجی ابزار مورد استفاده، یعنی دو رکن اساسی روایی و پایایی را در مرکز اصلی فعالیت‌های تحقیقاتی خود قرار دهند، چرا که نادیده گرفتن چنین موضوع مهمی، ممکن است به هدر رفتن تمام تلاش‌ها منجر شده و ضمن اتلاف غیرضروری منابع مالی، توصیه‌ها و اطلاعات نامعتبر در اختیار سیاست‌گزاران و برنامه‌ریزان قرار دهند (۱) و این نقص می‌تواند شانس بکارگیری یافته‌های تحقیقات را، توسط سایر پژوهشگران کاهش دهد (۵، ۶). هدف از تدوین این مقاله، بیان مراحل روش‌شناسی ارزیابی روایی (با تأکید بر روایی

بخش بسیار مهمی از تحقیقات مرتبط با حوزه‌های مختلف علوم اجتماعی، رفتاری، آموزشی و سلامت و به خصوص جمله آموزش پزشکی، مطالعه و بررسی مفاهیم انتزاعی و یا رفتارهای انسان است. در چنین بستری از تحقیقات، موضوع سنجش و اندازه‌گیری صفات و خصوصیات مورد مطالعه از جایگاه مهمی برخوردار خواهد شد. یکی از رایج‌ترین و متداول‌ترین روش‌های اندازه‌گیری چنین خصوصیاتی، استفاده از مقیاس‌ها، آزمون‌ها و پرسشنامه‌های محقق‌ساخته می‌باشد. بنابراین پژوهشگران و متخصصین به طور مستمر اقدام به طراحی و تطبیق مقیاس‌های اندازه‌گیری مورد استفاده با اهداف تحقیق و سنجش ویژگی‌های مورد مطالعه (مانند متغیرهای شناختی، رفتاری، احساسی و روانشناختی) در گروه‌های هدف، می‌نمایند (۱). ابزارها یا مقیاس‌ها ممکن است به منظور سنجش مفاهیم انتزاعی مانند دانش، ارزش‌های عاطفی و نگرشی، مهارت‌های روانی حرکتی یا شبیه‌سازی‌های بالینی به کار روند و یا به عنوان پرسشنامه‌های نظرسنجی جمعیتی مورد استفاده قرار گیرند. آموزش‌دهندگان علوم پزشکی به منظور ارتقاء صحت ارزیابی برنامه‌های آموزشی خود، به تلاش در زمینه‌ی تدوین و طراحی آزمون‌ها و پرسشنامه‌های برخوردار از روایی و پایایی مناسب اقدام می‌کنند (۲). از آنجا که بیشتر تحقیقات حوزه‌های آموزشی، روانشناختی و رفتاری با چنین رویکردی سروکار دارند، موضوع اعتبار و اعتماد قابل قبول ابزارهای سنجش و اندازه‌گیری، اهمیت پیدا می‌کند (۳). نتایج حاصل از هر تحقیقی باید تا آنجا که امکان دارد، قابل اعتماد باشد و هر مطالعه و تحقیقی باید در ارتباط با فرایندی که جهت تولید نتایج طی نموده است، مورد ارزیابی قرار گیرد (۴) و بر این اساس روایی و پایایی، دو جزء اساسی و ضروری مرتبط با یک ابزار سنجش محسوب می‌شوند (۲). اگرچه پژوهشگران

محتوا) و پایایی ابزار تحقیق، می‌باشد و ضمن بیان توالی روش شناسی فرایند مذکور، نحوه‌ی محاسبه شاخص‌های کمی و تفسیر نتایج نیز مورد بحث و تحلیل قرار گرفته است.

روایی چیست؟

روایی (Validity) به این موضوع اشاره می‌نماید که تفاسیر مرتبط با نمرات اخذ شده از آزمون‌های مورد استفاده برای یک منظور خاص، تا چه اندازه با شواهد و نظریه‌های علمی، مورد تأیید قرار می‌گیرد. به عبارت دیگر، روایی بیان می‌کند چگونه می‌توان نتایج یک آزمون را برای یک هدف خاص به درستی تفسیر نمود. تفسیر معنی نتایج سنجش حاصل از ابزارهای مختص کمیت‌های فیزیکی، مانند قد یا فشار خون، به سادگی و بطور مستقیم قابل دستیابی است، اما درک مفاهیم انتزاعی و پیچیده مانند آگاهی و یا نگرش مستقیمی و با چنین سهولتی غیرممکن است. در نتیجه محققین تلاش می‌کنند بطور نسبی و بر اساس یک ساختار پایه‌ای، مجموعه‌ای از مفاهیم انتزاعی و غیر قابل فهم را مورد ارزیابی قرار دهند. نتایج حاصل از یک ارزیابی روانسنجی، صرفاً در زمینه یا بستری که آن سازه مورد ارزیابی قرار می‌گیرد، معنی خواهد داشت (۷). روایی اساساً با صحت یا اعتبار نمرات اخذ شده از یک مقیاس اندازه‌گیری مرتبط بوده و به‌طور سنتی به‌عنوان «تثلیث مقدس» شناخته شده و شامل سه ویژگی مرتبط و در عین حال متمایز روایی محتوا، روایی سازه و روایی مرتبط با معیار می‌باشد (۱).

روایی محتوا یکی از معروف‌ترین و متداول‌ترین روش‌های ارزیابی اعتبار ابزارهای محقق ساخته می‌باشد و معمولاً در گام اول طراحی ابزار، مورد توجه پژوهشگران قرار می‌گیرد و شامل دو بخش بسیار مهم و ضروری و با دو رویکرد کیفی و کمی و با تمرکز بر روی دو گروه اصلی مورد مطالعه یعنی گروه هدف و گروه خبرگان مورد استفاده قرار می‌گیرد. روایی صوری (Face Validity) با رویکردی کیفی و به عنوان گام نخست ارزیابی روایی ابزار، یک نوع قضاوت عینی در مورد

ساختار ابزار محسوب می‌شود و به میزان یا درجه منطقی بودن یک آزمون، از نظر کسانی که به آن جواب می‌دهند، اشاره می‌کند و محقق تلاش می‌کند تا با انجام آن، نسبت به مرتبط بودن ابزار از نظر ظاهری با هدف مطالعه، مشابهت برداشت گروه هدف از گویه‌ها با نظرات محقق، موافقت گروه هدف با جمله‌بندی گویه‌های ابزار و قابل‌پذیرش بودن اجزا و کلیت ابزار از نظر گروه هدف (Target Group) اطمینان حاصل نماید (۳). با ارزیابی روایی محتوا (Content Validity) محقق تلاش می‌کند تا شواهد معتبر جهت کسب اطمینان از، لحاظ نمودن تمامی جوانب مهم و اصلی مفهوم مورد سنجش، سنجش صحیح همان مفهومی که قصد سنجش آن را دارد و قابلیت پذیرش اجزا و کلیت ابزار را، توسط متخصصین (Experts Panel) فراهم نماید (۸). در چنین فرایندی از ارزیابی اعتبار ابزار تحقیق، انجام روایی صوری و روایی محتوایی ابزار، به هر دو روش کمی و کیفی، حائز اهمیت اساسی است و در هر حال محوریت و تمرکز طراحی ابزار، باید بر روی گروه هدف قرار داشته باشد.

در بررسی روش‌شناسی مطالعات انجام شده در ایران با هدف ارزیابی مفاهیم انتزاعی مانند آگاهی و نگرش و یا ارزیابی تأثیر مداخلات آموزشی بر سطح آگاهی و نحوه‌ی نگرش گروه‌های هدف، به‌وضوح می‌توان پی‌برد که محققین در گزارش صحیح فرایند ارزیابی روایی ابزار تحقیق با مشکل روبه‌رو بوده و اطلاعات مرتبط با آن را ارایه نمی‌کنند و یا به ارایه‌ی اطلاعات ناکافی و مبهم بسنده می‌شود. در بسیاری از مطالعات حتی به تعداد دقیق اعضای پانل خبرگان اشاره نمی‌شود و یا تعدادی کمتر از پنج متخصص وارد مطالعه می‌شوند و بدون ارایه هرگونه توضیحی در باره‌ی تخصص و تجارب خبرگان، و بیان شاخص‌های کمی و یا فرایند کیفی ارزیابی روایی ابزار، به‌ذکر این عبارت مبهم یعنی «تأیید روایی ابزار توسط چند متخصص» بسنده می‌شود و در اکثر مواقع نیز نظرخواهی از گروه هدف به فراموشی سپرده می‌شود (۵، ۹،

گزینه‌های ۴ و ۵ را انتخاب نموده‌اند (جدول ۱)، مشخص (Frequency) و سپس مجموع نمرات اختصاص داده شده به هر گویه، به طور جداگانه محاسبه و میانگین نمرات هر گویه (Importance) تعیین شود. سپس با استفاده از فرمول: $Impact\ Item = Frequency(\%) \times Importance$ مقدار تأثیر هر گویه، محاسبه شده و چنانچه مقدار شاخص مذکور برای گویه‌ها بیش از ۱/۵ بود، آن گویه از نظر گروه هدف مهم تلقی شده و جهت انجام مراحل بعدی فرایند روانسنجی، مناسب ارزیابی شده و حفظ خواهد شد و در غیراینصورت باید حذف شود. در مرحله ارزیابی روایی ابزار به روش کیفی، نسبت به بررسی وجود مشکل در درک عبارات و کلمات، تناسب و ارتباط مطلوب گویه‌ها با یکدیگر، احتمال وجود ابهام و برداشت‌های نادرست از عبارات و یا وجود نارسایی در معانی کلمات، اقدام می‌شود و به این منظور از نمونه‌ی کوچکی از اعضای گروه هدف خواسته می‌شود تا گویه‌هایی را که از نظر آن‌ها نامناسب، غیرقابل فهم و یا دشوار هستند، مشخص کنند (۱۲). انتخاب تعداد حداقل ۱۰ نمونه با ویژگی‌های جمعیت شناختی مشابه جمعیت هدف اصلی، جهت محاسبه‌ی این شاخص مناسب به نظر می‌رسد (۱۰، ۱۳).

۱۰)، درحالی‌که حداقل تعداد اعضای پانل خبرگان مورد نیاز، جهت ارزیابی روایی ابزار تحقیق و امکان‌پذیر شدن محاسبه‌ی نسبت روایی محتوا پنج نفر است (۱۱). در بسیاری از موارد نیز محقق، در هنگام گزارش اعتبار یا روایی ابزار مورد استفاده در تحقیق، به مطالعات انجام شده توسط افراد دیگر استناد می‌کند، درحالی‌که در مطالعه مورد استناد، نیز ممکن است همچنان کیفیت و چگونگی فرایند روایی ابزار مورد استناد، با ابهام همراه باشد.

روایی محتوایی

ارزیابی روایی ابزار با تمرکز بر گروه هدف

این مرحله از ارزیابی بر روی گروه هدف متمرکز خواهد شد و به هر دو روش کمی و کیفی انجام می‌شود. جهت تعیین روایی ابزار به روش کمی، لازم است گویه‌های ابزار در قالب یک جدول با مقیاس لیکرت پنج درجه‌ای شامل: کاملاً مهم است (امتیاز ۵)، مهم است (امتیاز ۴)، به طور متوسط مهم است (امتیاز ۳)، اندکی مهم است (امتیاز ۲) و اصلاً مهم نیست (امتیاز ۱)، تدوین شده و در اختیار تعداد مناسب از افراد گروه هدف قرار داده‌شود تا نظرات خود را در مورد میزان اهمیت هر کدام از گویه‌های ابزار مشخص کنند. جهت محاسبه امتیاز تأثیر آیت، ابتدا باید نسبت افرادی که

جدول ۱: مقیاس لیکرت پنج درجه‌ای جهت اخذ نظرات گروه هدف در خصوص میزان اهمیت گویه‌های ابزار

شماره	گویه‌های ابزار	کاملاً مهم ۵	مهم ۴	بطور متوسط مهم ۳	اندکی مهم ۲	اصلاً مهم نیست ۱	امتیاز
۱							
۲							
۳							
...							

ارزیابی روایی محتوا با بهره‌گیری از پانل خبرگان: جهت ارزیابی روایی محتوا با بهره‌گیری از نظرات اعضای پانل خبرگان، تک‌تک گویه‌ها به دو روش کمی و کیفی مورد نظرخواهی قرار خواهد گرفت که در ذیل به توضیح آن‌ها خواهیم پرداخت:

نسبت روایی محتوا (Content Validity Ratio (CVR):

نسبت روایی محتوا یک آیتام آماری مفید برای رد و یا پذیرش گویه‌های ابزار است و در سطح بین‌المللی به عنوان روشی جهت ارزیابی روایی محتوا شناخته شده است (۱۴). این شاخص برای اولین بار در سال ۱۹۷۵ توسط Lawshe معرفی شد و به منظور حصول اطمینان از انتخاب مهم‌ترین و صحیح‌ترین محتوا (ضرورت گویه) بر اساس نظرخواهی از پانل خبرگان، مورد استفاده قرار گرفت (۱۱). حداقل تعداد مورد نیاز اعضای پانل خبرگان ۵ نفر می‌باشد و این تعداد می‌تواند تا ۴۰ نفر هم افزایش یابد. اعضای پانل خبرگان باید افرادی باشند که درباره موضوع تحقیق تخصص داشته و به عنوان کارشناس در آن زمینه شناخته شده، باشند. ایده‌آل این است که از طیفی متنوع از تخصص‌های مرتبط با

موضوع تحقیق استفاده شود. به‌طور معمول یک گروه پانل ۵ تا ۱۰ نفره می‌تواند مناسب باشد و استفاده از بیش از ده نفر ممکن است غیرضروری باشد. برای این منظور لازم است گویه‌های ابزار در قالب یک جدول طراحی و به‌طور جداگانه، در اختیار هر یک از اعضای پانل خبرگان قرار داده شود تا نظرات آن‌ها در خصوص ضرورت قرار گرفتن گویه‌ها در ابزار تحقیق، اخذ شود (۱۴). برای هر گویه یکی از سه طیف «۱. گویه ضروری است، ۲. مفید است اما ضروری نیست و ۳. ضرورتی ندارد»، پیش‌بینی شده و اعضای پانل خبرگان نظرات خود را در مورد از هر کدام از گویه‌ها، در فرم مربوطه ثبت خواهند نمود. پس از استخراج نظرات اخذ شده توسط افراد صاحب‌نظر، مقدار نسبت روایی محتوا با استفاده از فرمول (۱۴) محاسبه و مقدار عدد هرگویه، با مقادیر ارایه شده در جدول Lawshe و متناسب با تعداد اعضای پانل خبرگان، مورد مقایسه و ارزیابی قرار می‌گیرد و در صورتی که عدد نسبت روایی محاسبه شده، برابر و یا بیشتر از مقدار ارایه شده در جدول لاوشه بود، گویه‌ی مذکور حفظ و در غیر اینصورت باید از لیست گویه‌های ابزار حذف شود.

جدول ۲: نظرخواهی از اعضای پانل خبرگان در خصوص میزان ضرورت گویه‌های ابزار

شماره	گویه‌های ابزار	غیر ضروری	غیر ضروری مفید ولی غیر ضروری	ضروری
۱				
۲				
۳				
...				

جدول ۳: حداقل مقدار قابل پذیرش نسبت روایی محتوا بر اساس معیار لاوشه (۱۱)

مقدار CVR	تعداد اعضای صاحب نظر	مقدار CVR	تعداد اعضای صاحب نظر
۰/۵۱	۱۴	۰/۹۹	۵ الی ۷ نفر
۰/۴۹	۱۵	۰/۷۵	۸
۰/۴۲	۲۰	۰/۷۸	۹
۰/۳۷	۲۵	۰/۶۲	۱۰
۰/۳۳	۳۰	۰/۵۹	۱۱
۰/۳۱	۳۵	۰/۵۶	۱۲
۰/۲۹	۴۰	۰/۵۴	۱۳

استخراج داده‌ها اقدام شده و سپس مقدار شاخص روایی محتوا با استفاده از فرمول و برای هر کدام از سه ویژگی مذکور به طور جداگانه محاسبه شود و در انتها با محاسبه میانگین مقادیر محاسبه شده جهت هر گویه، مقدار کل شاخص روایی محتوایی، برای آن گویه محاسبه شود.

$$CVI = \frac{\text{تعداد متخصصینی که به گویه نمره 3 و 4 داده اند}}{\text{تعداد کل متخصصین}}$$

پس از محاسبه مقدار شاخص روایی محتوا جهت تمامی گویه‌ها، پذیرش هر کدام از آن‌ها براساس معیار زیر مورد تصمیم‌گیری قرار خواهد گرفت: «گویه‌های با نمرات بالاتر از ۰/۷۹ مناسب»، «گویه‌های با نمرات مابین ۰/۷۹ - ۰/۷۰ سوال برانگیز و نیازمند بازنگری و اصلاح»، «گویه‌های با نمرات کمتر از ۰/۷۰ غیر قابل قبول». گویه‌های مناسب و قابل پذیرش در ابزار، حفظ و گویه‌های غیرقابل قبول از لیست ابزار حذف خواهند شد. گویه‌های سوال برانگیز نیز، باید مورد بازنگری و بازنویسی قرار گیرند و نسبت به ارزیابی مجدد آن توسط پانل خبرگان اقدام شود.

شاخص روایی محتوا (CVI یا Content Validity Index): نسبت روایی محتوا، اطلاعات مرتبط با روایی آیتم‌ها را به طور جداگانه در اختیار قرار می‌دهد و اگر بخواهیم نسبت به روایی محتوایی کل ابزار آگاهی یابیم، می‌توان شاخص روایی محتوا را محاسبه نمود. شاخص روایی محتوا، به سادگی نمایانگر میانگین مقادیر نسبت روایی محتوا، برای تمامی گویه‌هایی است که مقدار نسبت روایی آن‌ها حداقل برابر ۰/۷۹ بوده و در ابزار نهایی حفظ شده‌اند (۱۴). Tilden و همکاران مقدار شاخص روایی محتوای مناسب برای پذیرش گویه را بالای ۰/۷۰ پیشنهاد می‌کنند (۱۵)، با وجود این Davis به مقادیر بالاتر از ۰/۸۰ به عنوان مقدار قابل پذیرش گویه تأکید می‌نماید (۱۶). شاخص روایی محتوا جهت اطمینان از اینکه گویه‌های ابزار به بهترین نحو جهت اندازه‌گیری محتوا طراحی شده‌اند، مورد استفاده قرار می‌گیرد. به این منظور تمامی گویه‌های ابزار باید در قالب یک جدول طراحی شود و به طور مجزا در اختیار تک تک اعضای پانل خبرگان قرار گیرد، تا نظرات خود را در مورد سه ویژگی مرتبط بودن، سادگی و وضوح تک تک گویه‌ها، با استفاده از یک طیف لیکرت چهار درجه‌ای، مشخص نمایند (۱۷، ۸). پس از اخذ نظرات اعضای پانل خبرگان، باید نسبت به

جدول ۴: نظرخواهی از اعضای پانل خبرگان در مورد ارتباط، سادگی و وضوح گویه‌های ابزار

شاخص روایی محتوا (CVI)														شماره		
میانگین کل شاخص روایی محتوا	وضوح				سادگی				مرتبط بودن						گویه ابزار	
	مقدار محاسبه شده	کاملاً واضح	واضح	ناحدردی واضح	غیر واضح	مقدار محاسبه شده	کاملاً ساده	ساده	ناحدردی پیچیده	پیچیده	مقدار محاسبه شده	کاملاً مرتبط	مرتبط			غیر مرتبط
																۱
																۲
																۳
																...

توالی مناسب و منطقی و از نظر فونت و قلم، زیبا طراحی شود و توسط گروه هدف به سهولت و بدون وجود هرگونه ابهامی، قابل خواندن و پاسخ‌گویی باشد.

پایایی چیست؟

در مطالعات حوزه‌ی روانشناختی و آموزشی، که یک رفتار یا ویژگی انتزاعی مورد سنجش قرار می‌گیرد، مبحث پایایی (Reliability) ابزار، همواره به عنوان یک نگرانی عمده مطرح است. یک ابزار اندازه‌گیری نمی‌تواند روا باشد، مگر اینکه از پایایی مناسب هم برخوردار باشد (۲). پایایی بیان می‌کند که اندازه‌گیری‌ها تا چه میزان قابل تکرار هستند و هرگونه اثر تصادفی که سبب گرایش اندازه‌گیری‌ها به سمت متفاوت شدن از موقعیتی به موقعیت دیگر شود، منبع خطای اندازه‌گیری محسوب می‌شود» (۱۸). Joppe مفهوم پایایی یا تکرارپذیری (Repeatability) را اینگونه بیان می‌کند: میزان برخورداری نتایج از ثبات در طی گذر زمان و چنانچه نتایج مطالعه با روش شناختی مشابه مجدداً تکرار شود، در اینصورت می‌توان ابزار تحقیق را پایا یا قابل اعتماد محسوب نمود (۱۹). Kirk و Miller در مطالعات کمی به سه نوع پایایی اشاره

ارزیابی روایی محتوا به روش کیفی:

در ادامه‌ی ارزیابی روایی ابزار به روش ارزیابی محتوایی، لازم است گویه‌های ابزار به لحاظ کیفی نیز توسط اعضای پانل خبرگان، مورد مطالعه و ارزیابی قرار گیرند. به این منظور لازم است از اعضای پانل خبرگان درخواست شود تا نظرات و پیشنهادات اصلاحی مورد نظرشان را، در مورد تک‌تک گویه‌ها از نظر: مفهوم و محتوای گویه‌ها، رعایت دستور زبان فارسی، تعداد کلمات و طول جملات، ترتیب و توالی گویه‌ها، پیشنهاد افزوده شدن گویه‌های جدید، وجود تناسب با ویژگی‌های فرهنگی و اجتماعی گروه هدف، و ساختار کلی ابزار، به صورت کتبی و در صورت لزوم به‌طور مبسوط بیان نمایند. پس از اخذ نظرات و پیشنهادات اعضای پانل خبرگان، گروه تحقیق باید نسبت به جمع‌بندی و اعمال نظرات و پیشنهادات و انجام اصلاحات ضروری در گویه‌های پرسشنامه اقدام نمایند. در صورتی‌که در گویه‌های ابزار به لحاظ تعداد و یا محتوای گویه‌ها تغییرات اساسی ایجاد شد، ممکن است تکرار مجدد فرایند ارزیابی روایی ابزار توسط گروه هدف و پانل خبرگان ضرورت یابد. به‌هرحال ضروری است گویه‌های مندرج در ابزار تحقیق، تا آنجا که ممکن است ساده و گویا، با

درونی، همگنی و یا تک بعدی بودن می تواند به بهبود استفاده از آلفا کمک کند (۲). ضروری است همبستگی درونی گویه های ابزار، قبل از بکارگیری آن در تحقیق، مورد ارزیابی قرار گیرد و چنانچه ابزار تحقیق، از چند بعد یا سازه تشکیل شده است، باید نسبت به وجود همبستگی درونی مناسب میان گویه های هر بعد نیز اطمینان حاصل نمود. تعداد گویه های ابزار، همبستگی آن ها با هم و ابعاد ابزار، مقدار ضریب شاخص آلفا را تحت تأثیر قرار می دهد. گزارش های متعددی در خصوص مقادیر قابل قبول آلفا وجود دارد و مقادیر ۰/۹ و بیشتر عالی، مقادیر مابین ۰/۸ تا ۰/۹ خوب، مقادیر مابین ۰/۷ تا ۰/۸ قابل قبول، مقادیر مابین ۰/۵ و ۰/۶ سؤال برانگیز و مقادیر کمتر از ۰/۵ غیر قابل قبول تلقی می شود. مقادیر پایین آلفا، ممکن است از تعداد کم گویه های ابزار، همبستگی ضعیف مابین گویه ها و یا سازه های ناهمگون نشأت گرفته باشد (۲).

آزمون-بازآزمون (Test-Retest):

چارلز Charles معتقد است که مفهوم پایایی یا ثبات، زمانی تحقق می یابد که پاسخ های ارایه شده به گویه های پرسشنامه و یا نمرات پاسخگوها در صورت انجام آزمون-بازآزمون، در طی دو دوره ی زمانی متفاوت، نسبتاً ثابت باقی بماند. به عبارت دیگر چنانچه یک صفت را با یک ابزار ثابت بسنجیم، نتایج اخذ شده باید مشابه باشند. میزان بالایی از ثبات، نشان دهنده ی میزان بالایی از پایایی است و مفهوم چنین چیزی، تکرارپذیری نتایج می باشد. البته روش آزمون-بازآزمون، خود می تواند سبب کاهش پایایی ابزار شود، چرا که پاسخگویی مجدد به گویه ها ممکن است باعث ایجاد حساسیت در آزمودنی شده و این موضوع ممکن است پاسخ های ارایه شده توسط وی در مرحله بازآزمون را تحت تأثیر قرار دهند (۲۱). بنابراین نمی توان اطمینان کامل یافت که تحت تأثیر چنین عوامل خارجی، هیچگونه تغییری در نگرش افراد رخ نخواهد داد. در نتیجه این موضوع می تواند پاسخ های آزمودنی ها را دستخوش تغییرات نماید. همچنین در صورت تکرار

می کنند که عبارتند از: درجه ی اخذ نتایج مشابه در صورت تکرار سنجش، ثبات مقادیر اندازه گیری شده در طول گذر زمان، تشابه اندازه گیری ها در خلال یک دوره ی زمانی معین (۲۰، ۲۱). پایایی ابزار را می توان با در نظر گرفتن اصولی از قبیل نگارش واضح و شفاف گویه ها، استفاده از دستورالعمل های تکمیل آزمون که به آسانی توسط پاسخگوها قابل درک باشند و آموزش مؤثر ممتحنین در زمینه ی نمره دهی شفاف و دقیق، بهبود بخشید (۳). اگرچه جهت ارزیابی پایایی ابزار تحقیق روش های گوناگونی وجود دارد، اما دو نوع از مداول ترین آن ها عبارتند از بررسی همبستگی درونی و روش آزمون-بازآزمون که در این مقاله مورد اشاره قرار می گیرند.

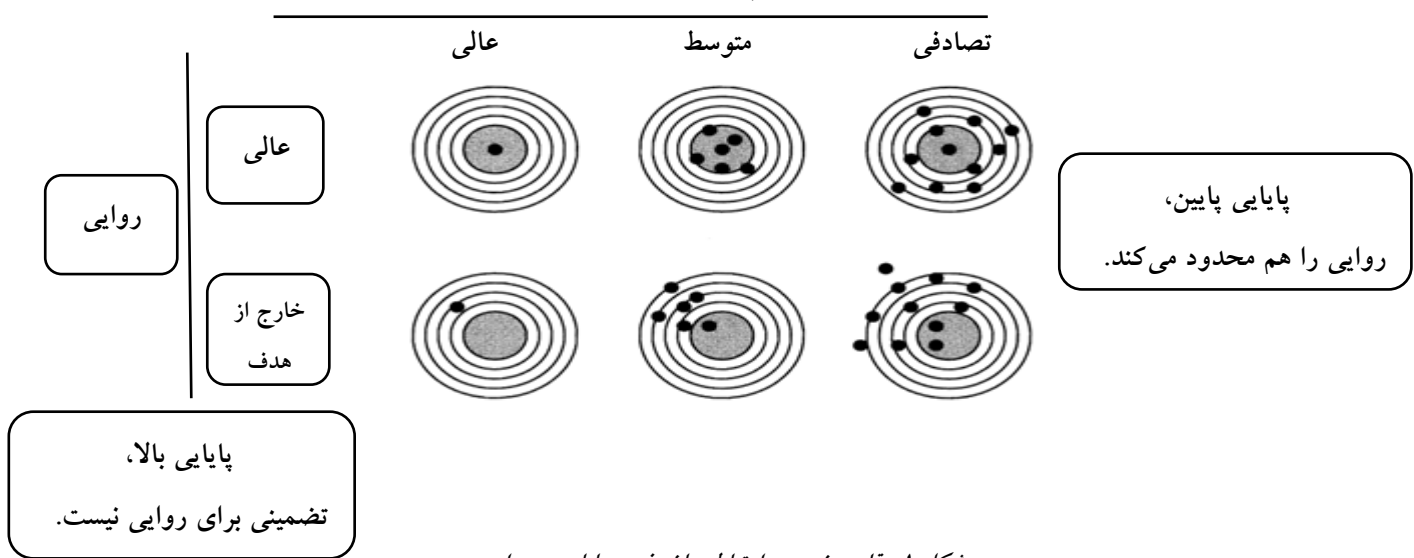
همبستگی درونی گویه ها:

بدون شک ضریب آلفای کرونباخ (Cronbach's alpha) یکی از مهم ترین و فراگیرترین شاخص های آماری در تحقیقات مرتبط با طراحی و بکارگیری آزمون ها محسوب می شود و در بسیاری از حوزه های علوم از جمله روان شناسی، آموزشی، اجتماعی، جامعه شناسی، اقتصادی، سیاسی، حقوق، و بخصوص در آموزش علوم پزشکی و پرستاری، مورد استفاده قرار می گیرد (۲، ۱۸)، چرا که در مقایسه با سایر روش های ارزیابی پایایی، کاربرد آسان تری دارد و فقط به یک بار اندازه گیری نیاز دارد (۲). علی رغم بکارگیری بسیار گسترده ی روش آلفا در متون علمی، کاربرد و تفسیر مناسب آن به درستی درک نشده است (۲، ۲۲). این شاخص در سال ۱۹۵۱ توسط Lee Cronbach و به منظور سنجش همبستگی درونی گویه های یک آزمون یا مقیاس معرفی شد و نمرات مابین ۰ و ۱ را در بر می گیرد (۲). این شاخص بیان می کند که آیا تمامی گویه های مقیاس طراحی شده، از همبستگی مناسب با مفهوم مورد سنجش برخوردار هستند یا خیر؟ استفاده نادرست از آلفا ممکن است به کنار گذاشتن نابجای آزمون و یا انجام تفسیرهای ناصحیح از نتایج تحقیق منجر شود. جهت اجتناب از چنین وضعیتی درک مناسب ارتباط مفاهیمی از قبیل ثبات

در شکل مذکور مشاهده می‌شود، پایایی به معنای قابل تکرار بودن نمرات مشابه، و روایی به معنی دقیقاً قرار گرفتن در نقطه‌ی مرکزی یک هدف خاص می‌باشد. هم در قسمت بالا و هم در قسمت پایین شکل ۱، با حرکت از سمت چپ به سوی سمت راست، از میزان پایایی کاسته می‌شود و پیامد چنین تغییری، تحت تأثیر قرار دادن روایی است و همانگونه که دایره‌ی واقع در سمت راست و بالای شکل ۱ نشان می‌دهد، با کاهش پایایی، روایی نیز به طور فزاینده‌ای به یک رخداد شانسی یا تصادفی تبدیل می‌شود. در عین حال دایره سمت چپ واقع در قسمت پایین شکل ۱، نشان از پایایی بسیار بالا می‌دهد، اما بالا بودن پایایی به هیچ وجه نمی‌تواند روایی بالا را تضمین کند (۲۴). بنابراین اگر ابزار مورد نظر از روایی مناسب برخوردار نباشد، در آن صورت پایایی بالا به تنهایی نیز، هیچ مفهوم و یا ارزشی نخواهد داشت. بنابراین بهترین حالت ممکن از نظر روایی و پایایی مناسب و قابل قبول بر اساس شکل ۱، دایره‌ی واقع در سمت چپ و بالای تصویر می‌باشد که هم در تکرار آزمایش، نمرات اخذ شده عیناً مشابه هم هستند و هم دقیقاً در نقطه‌ی مرکزی هدف قرار گرفته‌اند.

مجموعه‌ای از سوالات آزمون در زمان‌های متوالی، احتمال دارد برخی ویژگی‌های آزمودنی‌ها، بر انتخاب متفاوت پاسخ‌ها توسط آن‌ها در مقایسه با دفعات قبلی تأثیر گذاشته و در نتیجه پایایی ابزار دچار تغییر شود و این موضوع می‌تواند به کاهش صحت و ثبات ابزار و نمرات آزمون منجر شود. از این رو وظیفه‌ی جلب اعتماد دیگران در خصوص ثبات و دقت آزمون‌ها و نمرات حاصل از ابزارهای سنجش و اندازه‌گیری حاصل از یک تحقیق، بر دوش پژوهشگران قرار دارد. البته فاصله زمانی مابین آزمون اولیه و آزمون مجدد نیز مورد مناقشه می‌باشد (۲۳) و در مطالعات مختلف از چند ساعت تا ۶ ماه گزارش شده‌است، اما فاصله زمانی ما بین دو هفته تا یک ماه بیشتر مورد توافق همگانی قرار دارد. فاصله دو آزمون باید به اندازه کافی طولانی باشد تا آزمودنی، پاسخ‌های اولیه خود را به خاطر نیارد و نباید آنقدر طولانی باشد که محتواهای مرتبط با دانش و یا نگرش مرتبط با موضوع تحقیق دچار تغییر شود. به‌عنوان یک اصل، بیشتر شدن فاصله مابین دو آزمون، کمتر شدن پایایی ابزار را به همراه دارد (۲۳). شکل ۱، به درک بهتر مفهوم روایی و پایایی و اهمیت بسیار زیاد ارتباط میان این دو ویژگی کمک می‌کند. همان‌گونه که

پایایی



شکل ۱: مقایسه نحوه‌ی ارتباط میان مفهوم پایایی و روایی

اهمیت پایایی و ثبات ابزار آزمون نیز، هیچ شکی وجود ندارد، اما خصوصیت روایی ابزار، هم به لحاظ اهمیت و هم از نظر تقدم، بر پایایی ابزار اولویت دارد. به عبارت دیگر چنانچه ابزار تحقیق، از روایی مناسب و قابل قبول برخوردار نباشد، یعنی قادر به سنجش درست و دقیق آنچه که باید بسنجد، نباشد، آن وقت گفتگو در مورد نحوه‌ی ارزیابی و کیفیت پایایی آن، چه معنا و مفهومی خواهد داشت؟

بدون شک هر مطالعه‌ای ممکن است از محدودیت‌ها و یا نواقصی برخوردار باشد، اما واضح است که اعتبار یک مطالعه در درجه‌ی اول، وابسته به روش‌شناسی تحقیق است. اگرچه تمام اجزای روش‌شناسی تحقیق، حائز اهمیت خاص خود هستند، اما چنانچه بخواهیم فقط یکی از آن‌ها را، به عنوان مهم‌ترین جزء انتخاب کنیم، به احتمال بسیار زیاد آن انتخاب، روش‌ها و ابزارهای سنجش متغیرهای تحقیق (اعم از کمی و یا کیفی) خواهد بود، چراکه طراحی یک مطالعه، با هر میزانی از دقت و اطمینان هم انجام شود، ولی چنانچه ابزار سنجش متغیرهای مطالعه، فاقد اعتبار و اعتماد کافی باشد، آن وقت امکان اظهار نظر در خصوص نتایج تحقیق، میزان دستیابی به اهداف و همینطور پذیرش یا عدم پذیرش فرضیه‌های تحقیق وجود نخواهد داشت و در چنین وضعیتی، بیان هرگونه بحث و تفسیر در باره‌ی نتایج تحقیق و یا مقایسه‌ی آن با سایر مطالعات انجام شده، مخدوش و بی‌اعتبار خواهد بود.

در این مطالعه تلاش شد ضمن بیان اهمیت و ضرورت دو خصوصیت بسیار اساسی ابزارهای سنجش و اندازه‌گیری یعنی روایی و پایایی، به‌طور ویژه نسبت به بیان اهمیت و مراحل انجام ارزیابی روایی محتوا، به صورت گام به گام و همراه با جزئیات آن، و همینطور دو نوع از متداول‌ترین روش‌های ارزیابی پایایی ابزار برداریم. انتظار می‌رود توجه به این‌گونه مباحث پایه در روش‌شناسی تحقیقات علوم پزشکی و سایر حوزه‌های پژوهشی، بیش‌ازپیش مد نظر پژوهشگران و

در مطالعات انجام شده در ایران در مقایسه با مبحث روایی ابزار، در گزارش پایایی ابزار مشکلات کمتری به چشم می‌خورد، ولی غالباً جهت ارزیابی پایایی ابزار به بررسی سازگاری درونی گویه‌ها و گزارش مقدار آلفای کرونباخ اکتفا می‌شود، در حالی‌که ثبات نتایج حاصل از تکرار آزمون‌ها کمتر مورد توجه قرار می‌گیرد. دشواری دسترسی مجدد به نمونه‌ها و یا کم اهمیت تلقی نمودن تکرار آزمون در یک دوره‌ی زمانی کوتاه از نظر محقق یا آزمودنی‌ها، شاید از دلایل استقبال کمتر محققین از انجام این روش ارزیابی پایایی ابزار باشد.

نتیجه‌گیری

مرور اجمالی روش‌شناسی مطالعات مرتبط با سنجش مفاهیم انتزاعی مانند آگاهی، نگرش و باورهای بهداشتی در ایران به وضوح نشان می‌دهد که پژوهشگران در گزارش صحیح فرایند ارزیابی روایی و پایایی ابزارهای تحقیق با مشکل مواجه هستند و بطور متداول یا چنین فرایندی طی نمی‌شود و یا اطلاعات کامل ارایه نمی‌شود و یا به‌ارایه‌ی ناقص اطلاعات اکتفا می‌شود. از طرف دیگر در بخش قابل توجهی از مطالعات، اخذ نظرات گروه‌هدف به عنوان اولین و مهم‌ترین گام در ارزیابی روایی ابزار مورد توجه قرار نمی‌گیرد و در مورد نظرخواهی از گروه صاحب‌نظران، نیز غالباً به‌صورت مبهم و بدون ذکر تعداد و تخصص خبرگان و ارایه‌ی مستندات مرتبط با شاخص‌های کمی و کیفی روانسنجی، صرفاً به بیان عبارت «تأیید روایی ابزار توسط چند متخصص و یا کمتر از پنج متخصص» اشاره می‌شود، در حالی‌که چنین شیوه‌ای فاقد اعتبار لازم می‌باشد. البته در مورد گزارش پایایی در مقایسه با مبحث روایی ابزار، این مشکل از شدت کمتری برخوردار است و غالباً با گزارش مقدار آلفای کرونباخ، وضعیت پایایی ابزار، مورد اشاره قرار می‌گیرد، اما معمولاً ارزیابی ثبات ابزار به روش آزمون- بازآزمون نادیده گرفته می‌شود. البته در

خارجی، کیفیت و چگونگی تأمین و ارزیابی مستندات مرتبط با ارزیابی روایی و پایایی ابزارهای سنجش و اندازه‌گیری مورد استفاده در تحقیق می‌باشد.

نویسندگان قرارگیرد، چرا که امروزه یکی از مهم‌ترین معیارها و ضوابط تأثیر گذار بر فرایند پذیرش و یا عدم پذیرش یافته‌های پژوهشی، در مجلات معتبر داخلی و

References

- 1- Barry AE, Chaney B, Piazza-Gardner AK, Chavarria EA. Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education & Behavior*. 2014;41(1):12-8.
- 2- Tavakol M, Dennick R. Making sense of Cronbach's alpha. *International journal of medical education*. 2011;2:53-5.
- 3- Drost EA. Validity and reliability in social science research. *Education Research and perspectives*. 2011;38(1):105-23.
- 4- Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today*. 2004; 24 (2):105-12.
- 5- Haghdoost A, Pourkhandani A, Motaghipisheh S, Farhoudi B, Fahimifar N, Sadeghirad B. Knowledge and Attitude concerning HIV/AIDS among Iranian Population: a Systematic Review and Meta- Analysis. *Iranian Journal of Epidemiology*. 2011;6(4):8-20.
- 6- Nejat S, Feyzzadeh A, Asghari S, Keshtkar A, HESHMAT R, Majdzadeh S. HIV risk factors in Iran; systematic review, meta-analysis and generalized impact fraction approaches. *PAYESH*. 2007;6(1):45-54.
- 7- Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*. 2006;119(2):166.e7-16.
- 8- 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.
- 9- Mohammadbeigi A, Mohammadsalehi N, Aligol M. Validity and Reliability of the Instruments and Types of MeasurmentS in Health Applied Researches. *Journal of Rafsanjan University of Medical Sciences*. 2015;13(12):1153-70.
- 10- Vakili MM, Hidarnia AR, Niknami S. Development and Psychometrics of an Interpersonal Communication Skills Scale (A.S.M.A) among Zanjan Health Volunteers. *Hayat*. 2012; 18 (1):5-19
- 11- Lawshe CH. A quantitative approach to content validity. *Personnel psychology*. 1975;28(4):563-75.
- 12- Lacasse Y, Godbout C, Series F. Health-related quality of life in obstructive sleep apnoea. *European Respiratory Journal*. 2002;19(3):499-503.
- 13- Vakili MM, Hidarnia AR, Niknami S,

- Mousavinasab N. Development and psychometrics of Health Belief Model instrument about HIV/AIDS. *Zahedan Journal of Research in Medical Sciences zahedan*. 2012;14(9):64-71.
- 14- Gilbert GE, Prion S. Making Sense of Methods and Measurement: Lawshe's Content Validity Index. *Clinical Simulation in Nursing*. 2016;12(12):530-1.
- 15- Tilden VP, Nelson CA, May BA. Use of qualitative methods to enhance content validity. *Nursing Research*. 1990;39(3):172-5.
- 16- Davis LL. Instrument review: Getting the most from a panel of experts. *Appl Nurs Res*. 1992;5(4):194-7.
- 17- 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.
- 18- Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol*. 1993;78(1):98-104.
- 19- Joppe M. The Research Process. Retrieved February 25, 1998. 2000.
- 20- Kirk J, Miller ML. Reliability and validity in qualitative research: Sage; 1986.
- 21- Golafshani N. Understanding reliability and validity in qualitative research. *The qualitative report*. 2003;8(4):597-606.
- 22- Schmitt N. Uses and abuses of coefficient alpha. *Psychol Assessment*. 1996;8(4):350-3.
- 23- DeVon HA, Block ME, Moyle- Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*. 2007;39(2):155-64.
- 24- Krippendorff K. Content Analysis An Introduction to Its Methodology. London: Sage Publications; 2004. P. 211-214.