

تعاریف و تفاوت های پایایی و روایی در پرسشنامه پژوهش

Reliability and Validity

نت روش تحقیق کارشناسی ارشد دانشگاه آزاد اسلامی واحد بین المللی جلفا

دکتر محمد حسن حدادی

حین نوشتن فصل سوم پایان نامه حتما به بخشی برواهید خورد که باید ابزار گردآوری اطلاعات را از نظر پایایی و روایی بسنجید.

می دانید که از ابزار مختلفی، که متداول ترین آن پرسشنامه است، برای گردآوری داده های فصل روش تحقیق استفاده می شود. اما این ابزار تا زمانی که دو ویژگی مهم پایایی و روایی را نداشته باشند قابل اعتماد نیستند و نمی توان به نتایج آن ها اطمینان کرد.

بنابراین لازم است ابتدا با این مفاهیم آشنا شوید و بعد بدانید ابزاری که برای گردآوری داده ها انتخاب کرده اید را چگونه از نظر این دو ویژگی بسنجید.

هیچ سنجشی بی خطا نیست. هر قدر هم که ابزار سنجش ما دقیق و هر قدر تلاش ما برای مشاهده دقیق باشد باز همواره سنجش آمیخته با خطاست. هر نوع سنجشی شامل درجه ای از خطاست که به آن " خطای سنجش " می گویند. اما هدف ما این است که خطای سنجش را تا حد امکان کاهش دهیم. مهمترین معیار درستی سنجش ، اعتبار آن است یعنی سنجش واقعا سنجش همان چیزی باشد که قصد سنجش آن را داریم. سنجش باید از ثبات هم برخوردار باشد یعنی هرگاه آن را دوباره تکرار کنند به همان نتایج برسند. این ثبات سنجش را پایایی می خوانند. البته وقتی سنجش از اعتبار برخوردار نباشد پایایی آن هم مناسبی نمی یابد. قبل از بکارگیری ابزارهای اندازه گیری لازم است پژوهشگر از طریق علمی، نسبت به روا بودن ابزار اندازه گیری مورد نظر و پایایی آن که مکمل هم بحساب می آیند اطمینان نسبی پیدا کند. ابزاری که برای جمع آوری داده ها مورد استفاده قرار می گیرد در مرحله نخست باید از روایی (اعتبار) برخوردار باشد و در مرحله دوم باید پایایی (اعتماد) داشته باشد. در واقع یکی از شروطی که باعث می شود یک تحقیق روایی داشته باشد این است که ابزار تحقیق، روایی داشته باشد.

مفهوم روایی (Validity) پاسخ دهنده این است که ابزار اندازه گیری تا چه حد خصیصه مورد نظر را می سنجد. موضوع روایی از آن جهت اهمیت دارد که اندازه گیری های نامتناسب می تواند هر پژوهش علمی را بی ارزش سازد.

مفهوم پایایی (Reliability) قابلیت تکرار روش یا ابزار اندازه گیری است. میزان سازگاری نتایج حاصل از اجرای مجدد آزمون با ابزار اندازه گیری را پایایی گویند. اگر روشی از پایایی برخوردار نباشد، داده های گردآوری شده روایی (اعتبار) نیز نخواهند داشت. در تحقیقات کمی به پایایی (Reliability) می گویند و در تحقیقات کیفی به آن Dependability می گویند.

منظور از روایی این است که مقیاس و محتوای ابزار یا سؤالات مندرج در ابزار دقیقاً متغیرها و موضوع مورد مطالعه را بسنجد؛ یعنی اینکه هم داده های گردآوری شده از طریق ابزار مزاد بر نیاز تحقیق نباشد و هم اینکه بخشی از داده های مورد نیاز در رابطه با سنجش متغیرها در محتوای ابزار حذف نشده باشد یا به عبارت دیگر، عین واقعیت را به خوبی نشان دهد (حافظ نیا، ۱۳۸۴: ۱۵۵).

مثال:

اگر بخواهیم شیوع کم وزنی هنگام تولد را در نوزادان یک زایشگاه بدانیم، باید همه نوزادان تولد یافته را وزن کنیم. برای این کار می بایست از ترازوی مخصوص توزین نوزاد و آنهم ترازوی استاندارد استفاده شود تا اهداف تحقیق را درست اندازه گیری کند و به عبارتی ابزار تحقیق از روایی برخوردار باشد. در این مثال سپس باید دید با چند بار وزن کردن آیا وزن یک کودک را در هر مرحله، مشابه وزن قبلی نمایش می دهد تا به پایایی دست یابیم.

روایی به چه معناست؟

ابزاری که برای جمع آوری داده ها مورد استفاده قرار می گیرند، در مرحله اول باید از روایی یا اعتبار (Validity) برخوردار باشند و در مرحله دوم باید پایایی یا اعتماد داشته باشند.

روایی بدین معناست که روش یا ابزار به کار رفته تا چه حد می تواند خصوصیت مورد نظر را درست اندازه گیری کند.

روایی: روایی عبارت است از درجه ای که آزمون یا اندازه گیری به مقدار واقعی نزدیک است. روایی یک آزمون یا وسیله اندازه گیری به روش های مختلفی قابل ارزیابی است. روایی همانند پایایی به وسیله ضریب همبستگی بیان می شود. یک مقدار منفی می تواند یک مقدار مثبت دارای مفهوم باشد. مثلاً یک شغل خاص ممکن است به قدرت فکری پایینی نیاز داشته باشد که این می تواند به طور غیر مستقیم (منفی) با آزمون ضریب هوشی بستگی داشته باشد. بنابراین کارگران مطلوب و مناسب برای این شغل افرادی هستند که دارای نمره پایینی در آن آزمون باشند.

مقصود از روایی این است که ابزاری که برای اندازه گیری مورد نظر انتخاب شده است ویژگی و خصوصیتی که ابزار برای آن طراحی شده است را دارد یا خیر. به عبارت دیگر مفهوم روایی، به این سوال پاسخ می دهد که ابزار اندازه گیری تا چه حد خصیصه مورد نظر را می سنجد.

اعتبار (روایی) به ارتباط منطقی، بین پرسش های آزمون و مطلب مورد سنجش اشاره دارد. وقتی گفته می شود آزمون، روایی دارد به این معنا است که پرسش های آزمون به طور دقیق آنچه را که مورد نظر می باشد، می سنجد. اعتبار، جنبه های مختلف دارد و ارتباط بین پرسش و آزمودنی با توجه به کلیه جنبه های آن حاصل می شود. در صورتی که این ارتباط وجود نداشته باشد اعتبار به وجود نمی آید.

علاوه بر اینکه روش اندازه گیری می بایست دارای پایایی و روایی بالایی باشد، باید عملی و در شرایط و با امکانات موجود قابل استفاده باشد. مثلاً یک روش اندازه گیری ممکن است در عین حالی که دارای بالاترین روایی است اما بسیار مشکل و دردسرساز باشد، بنابراین استفاده از آن به موارد ویژه ای محدود می گردد. با یک مثال این موضوع را بیشتر توضیح می دهیم.

فرض کنید از ابزار پرسشنامه برای ارزیابی «تعهد سازمانی کارکنان» استفاده می کنید ولی پرسش های آن طوری طراحی شده اند که به نظر می رسد «رضایت شغلی کارکنان» در حال ارزیابی است.

نتایج به دست آمده از این پرسش نامه روایی ندارد چرا که خصیصه ی مورد نظر که در اینجا «تعهد سازمانی» بوده ارزیابی نشده است.

بنابراین وقتی گفته می شود ابزار گردآوری داده ها باید روایی داشته باشند بدین معناست که چه پرسشنامه تهیه می کنید چه می خواهید مصاحبه کنید یا ارزیابی خود را با مشاهده حضوری انجام دهید باید چیزی که سنجیده می شود مرتبط با هدفی باشد که پیش گرفته اید در غیر این صورت نتیجه نهایی از ارزش خاصی برخوردار نخواهد بود.

یک اندازه گیری ممکن است از پایایی بالایی برخوردار باشد بدون اینکه روایی داشته باشد، اما روایی بالا همیشه نشان دهنده پایایی بالا نیز خواهد بود.

پایایی به چه معناست؟

پایایی یک معیار یا آزمون خود بایستی دارای پایایی باشد. پایایی عبارت است از درجه توافق دو اندازه گیری مستقل و جدا از هم از یک چیز. این توافق معمولاً به صورت ضریب همبستگی بین آن دو اندازه گیری می شود. (پایایی عبارت است از میزان همبستگی بین نتایج حاصله از دو بار اندازه گیری که به طور مستقل و

جداگانه بر روی متقاضیان صورت می گیرد). به کمک پایایی می توان به میزان خطا پی برد. در پایایی یک وسیله اندازه گیری یا یک آزمون به وسیله ضریب همبستگی بیان می گردد. علامت این ضریب (مثبت یا منفی) مشخص کننده جهت همبستگی است و مقدار آن که بین یک و صفر متغییر است را نشان می دهد. هر چه ضریب همبستگی بزرگتر باشد خطای اندازه گیری کمتر و دقت و صحت نتایج به دست آمده بیشتر خواهد بود.

اعتبار اندازه گیری (روایی) ابزار اندازه گیری

مفهوم اعتبار (روایی) به این سوال پاسخ می دهد که ابزار اندازه گیری تا چه حد خصیصه مورد نظر را می سنجد. بدون آگاهی از اعتبار ابزار اندازه گیری نمی توان به دقت داده های حاصل از آن اطمینان داشت. ابزار اندازه گیری ممکن است برای اندازه گیری یک خصیصه ویژه دارای اعتبار باشد، در حالی که برای سنجش همان خصیصه بر روی جامعه دیگر از هیچ گونه اعتباری برخوردار نباشد.

برای مثال یک آزمون ریاضی ممکن است برای سنجش توانایی ریاضی دانش آموزان پایه پنجم ابتدایی از اعتبار لازم برخوردار باشد اما برای سنجش توانایی ریاضی دانش آموزان پایه سوم راهنمایی فاقد اعتبار باشد. روش های متعددی برای تعیین اعتبار ابزار اندازه گیری وجود دارد که در این جا به اختصار در مورد هر یک توضیح داده خواهد شد؛

روایی یا اعتبار محتوی، روایی یا اعتبار صوری، روایی یا اعتبار ملاکی، روایی یا اعتبار پیش بین، روایی یا اعتبار همزمان، روایی یا اعتبار سازه، روایی یا اعتبار استقرایی، روایی یا اعتبار همخوان، روایی یا اعتبار وفاقی، روایی یا اعتبار همگرا، روایی یا اعتبار تعریفی، روایی یا اعتبار افتراقی، روایی یا اعتبار تفکیکی، روایی یا اعتبار تجربی، روایی یا اعتبار عاملی، روایی یا اعتبار نمونه گیری، روایی یا اعتبار وابسته به قوانین طبیعی، روایی یا اعتبار، ترکیبی، روایی یا اعتبار ویژگی و روایی یا اعتبار افزایشی.

رابطه بین روایی و پایایی

یک آزمون باید پایا باشد تا بتواند روا باشد. اگر آزمونی در هر بار اجرا بر روی تعدادی دانش آموز نتایج مختلفی را بدست بدهد آن آزمون پایا نخواهد بود و در واقع هیچ چیز را به درستی اندازه نخواهد گرفت و اگر یک آزمون چیزی را به درستی اندازه گیری نکند هیچ اطلاع مفیدی به ما نخواهد داد. پس برای اینکه یک آزمون روا باشد نخست پایا باشد اما روایی برای پایایی ضروری نیست .

شیوه های تعیین پایایی :

برای تعیین ضریب پایایی روشهای مختلفی وجود دارد :

- روش پایایی مصححان

- روش بازآزمایی یا آزمون مجدد test-retest
- روش تصنیف یا دو نیمه کردن آزمون یا همسانی درونی Split-half
- روش فرم های موازی یا آزمون های همتا Equivalence
- روش کودر-ریچاردسون Kuder-Richardson
- روش استفاده از مشخصه های آماری---آلفای کرونباخ

قابلیت اعتماد و اعتبار ابزار اندازه گیری (پایایی و روایی)

قابلیت اعتماد (پایایی یا Reliability) و همچنین اعتبار (روایی یا Validity) یک پرسشنامه یا ابزار اندازه گیری، از موضوعات بسیار مهم در امر جمع آوری اطلاعات و مشاهدات می باشد که متأسفانه گاهی مورد غفلت قرار می گیرد.. در اینجا به شرح این دو موضوع مهم می پردازیم:

۱- قابلیت اعتماد ابزار اندازه گیری

قابلیت اعتماد که واژه هایی مانند پایایی، ثبات و اعتبار برای آن به کار برده می شود و معادل انگلیسی آن Reliability است، یکی از ویژگی های ابزار اندازه گیری (پرسشنامه یا مصاحبه یا سایر آزمون های علوم اجتماعی) است. مفهوم یاد شده با این امر سروکار دارد که ابزار اندازه گیری در شرایط یکسان تا چه اندازه نتایج یکسانی به دست می دهد. از جمله تعریف هایی که برای قابلیت اعتماد ارائه شده است می توان به تعریف ارائه شده توسط ایبل و فریسبی (۱۹۸۹) اشاره کرد: «همبستگی میان یک مجموعه از نمرات و مجموعه دیگری از نمرات در یک آزمون معادل که به صورت مستقل بر یک گروه آزمودنی به دست آمده است.»

با توجه به این امر معمولاً دامنه ضریب قابلیت از صفر (عدم ارتباط) تا ۱+ (ارتباط کامل) است. ضریب قابلیت اعتماد نشانگر آن است که تا چه اندازه ابزار اندازه گیری ویژگی های با ثبات آزمودنی و یا ویژگی های متغیر و موقتی وی را می سنجد. قابلیت اعتماد در یک آزمون می تواند از موقعیتی به موقعیت دیگر و از گروهی به گروه دیگر متفاوت باشد. برای محاسبه ضریب قابلیت اعتماد ابزار اندازه گیری شیوه های مختلفی به کار برده می شود. از آن جمله می توان به الف) اجرای دوباره (روش بازآزمایی) ، ب) روش موازی (همتا) ، ج) روش تصنیف (دو نیمه کردن) ، د) روش کودر- ریچاردسون و سایر اشاره کرد.

روش آلفای کرونباخ

ضریب آلفای کرونباخ توسط کرونباخ در سال ۱۹۵۱ ابداع شده و یکی از متداولترین روشهای اندازه گیری اعتماد پذیری و یا پایایی (Reliability) پرسش نامه ها است. ضریب آلفای کرونباخ برای سنجش میزان تک بعدی بودن نگرشها، قضاوت ها، عقاید و سایر مقولاتی که اندازه گیری آنها آسان نیست به کار می رود.

در واقع می‌خواهیم ببینیم تا چه حد برداشت پاسخگویان از سؤالات یکسان بوده است. اساس ضریب آلفای کرونباخ نیز پایه طیف ها یا مقیاس هاست. مقیاس عبارت است از دسته ای از اعداد که بر روی یک پیوستار به افراد، اشیاء یا رفتار در جهت به کمیت کشاندن کیفیتها اختصاص داده می شود. وانگهی مقیاسها جهت اندازه گیری گرایش و نگرشها به کار می رود .

رایج ترین مقیاسی که در تحقیقات اجتماعی به کار برده می شود لیکرت است و آن مقیاسی است که بر سطح سنجش در مقیاس ترتیبی که دارای تعدادی گویه ، معمولا بین ۵ تا ۷ گویه می باشد که به صورت منفی و مثبت سازمان یافته و تعداد آنها با هم برابر است. به هر یک از گزینه ها نمراتی، مثلاً از ۱-۷ و یا ۰-۶ یا هر شکل دیگری که با ملاحظه مثبت و منفی بودن سؤال ترتیب نیز مراعات گردد داده می شود .

مثال: نظر شما در باره سیاست خودکفایی گندم که در سال سه دهه اخیر در ایران دنبال می شود چیست؟
کاملاً موافقم ... موافقم... نظری ندارم.. مخالفم.. کاملاً مخالفم..

چون فرمول آلفای کرونباخ مبتنی بر واریانس است، بنابراین طبق خواص واریانس به هر ترتیبی نمره داده می شود مقدار آن تغییر پیدا نمی کند .

هرقدر شاخص آلفای کرونباخ به ۱ نزدیکتر باشد، همبستگی درونی بین سؤالات بیشتر و در نتیجه پرسشها همگن تر خواهند بود. کرونباخ ضریب پایایی ۴۵٪ را کم، ۷۵٪ را متوسط و قابل قبول، و ضریب ۹۵٪ را زیاد پیشنهاد کرده (کرونباخ ۱۹۵۱). در بسیاری از منابع نیز مقادیر به دست آمده بالای ۰/۷ در این آزمون مطلوب تلقی می‌شود. بدیهی است در صورت پایین بودن مقدار آلفا، بایستی بررسی شود که با حذف کدام پرسشهای پرسشنامه مقدار آن را می توان افزایش داد.

آلفای کرونباخ با استفاده از رابطه زیر قابل محاسبه است :

در این رابطه K تعداد پرسشها و S_i انحراف معیار امتیاز کل پرسشها است .

هر قدر همبستگی مثبت بین سؤالات بیشتر شود، میزان آلفای کرونباخ بیشتر خواهد شد و بالعکس هر قدر واریانس میانگین سؤالات بیشتر شود آلفای کرونباخ کاهش پیدا خواهد کرد. افزایش تعداد سؤالات تاثیر مثبت و یا منفی (بسته به نوع همبستگی بین سؤالات) بر میزان آلفای کرونباخ خواهد گذاشت. افزایش حجم نمونه باعث کاهش واریانس میانگین سؤالات در نتیجه باعث افزایش آلفای کرونباخ خواهد شد.

لازم به ذکر است در صورتیکه پرسشنامه مورد بررسی دارای سؤالات ۵ گزینه ای و ۲ گزینه ای باشد، باید برای هر کدام از حالتها ضریب آلفای جداگانه ای محاسبه گردد، چرا که ممکن است برای سؤالات ۵ گزینه ای آلفا مورد قبول باشد ولی برای برای ۲ گزینه ای مورد قبول واقع نگردد.

برای محاسبه ضریب آلفای کرونباخ ابتدا باید واریانس نمره های هر زیر مجموعه سوال های پرسشنامه (یا زیر آزمون) و واریانس کل را محاسبه کرد. سپس با استفاده از فرمول زیر مقدار ضریب آلفا را محاسبه کرد.

$$r_a = \frac{j}{j-1} \left(1 - \frac{\sum S_j^2}{S^2} \right)$$

که در آن:

$$j = \text{تعداد زیر مجموعه سوال های پرسشنامه یا آزمون.}$$

$$S_j^2 = \text{واریانس زیر آزمون } j\text{ام.}$$

$$S^2 = \text{واریانس کل آزمون.}$$

برای مثال اگر پرسشنامه دارای سه زیر مجموعه سوال باشد، به طوری که واریانس هر زیر مجموعه به ترتیب ۶، ۴ و ۷ بوده و واریانس کل برابر با ۳۲ باشد، مقدار ضریب آلفای کرونباخ با استفاده از فرمول یاد شده به صورت زیر به دست می آید:

$$r_a = \frac{3}{2} \left(1 - \frac{17}{32} \right) = 0.8$$

مقدار صفر این ضریب نشان دهنده عدم قابلیت اعتماد و +۱ نشان دهنده قابلیت اعتماد کامل است.

روش کودر- ریچاردسون

۷ نت روش پژوهش کارشناسی ارشد، دکتر محمد حسن حدادی دانشگاه آزاد اسلامی واحد بین المللی جلفا

<http://www.kenareh-gostare-aras.ir/>

آزمون تنها یکبار اجرا می شود، اما در این روش همه ماده های آزمون تحلیل می شوند. کودر و ریچاردسون برای بررسی همسانی درونی آزمون و تعیین پایایی آن دو فرمول مورد استفاده قرار داده اند که به KR₂₀, KR₂₁ شهرت دارند .

عوامل موثر در اعتبار یک آزمون :

- افزایش تعداد سوالات یک آزمون باعث افزایش میزان اعتبار آن میشود .
- حذف سوالهای ناممکن باعث افزایش اعتبار آزمون میگردد .
- افزودن سوالهایی که از نظر دشواری در حد متوسط باشند میزان اعتبار بالا میرود .
- افزایش تعداد سوالهایی که افراد را از نظر ویژگیهای مورد نظر اندازه گیری و بخوبی از همدیگر متمایز سازد .

کودر و ریچاردسون دو فرمول را برای محاسبه هماهنگی درونی آزمون ها ارائه نموده اند. فرمول اول برآوردی از میانگین ضرایب قابلیت اعتماد برای تمام طرق ممکن تنصیف (دو نیمه کردن) را به دست می دهد. این ضریب برابر است با:

$$r_1 = \frac{n}{n-1} \left(1 - \frac{\sum pq}{S^2} \right)$$

که در آن :

n تعداد سوال ها - P نسبت پاسخ های صحیح - q نسبت پاسخ های غلط - S^2 واریانس نمرات کل

فرمول دوم: برای استفاده از فرمول دوم باید سطح دشواری سوال های آزمون یکسان باشد:

$$r_2 = \frac{n}{n-1} \left(1 - \frac{\bar{x}(n-\bar{x})}{nS^2} \right)$$

که در آن :

n تعداد سوال ها - S^2 واریانس نمره های کل - \bar{x} میانگین نمرات

برای محاسبه ضریب قابلیت اعتماد با این روش، نیازی به داشتن نسبت پاسخ های صحیح و غلط نیست، بلکه در دست داشتن واریانس نمره های کل و میانگین نمرات کفایت می کند.

برای مثال، آزمونی دارای ۴۰ سوال است. این آزمون به گروهی از دانش آموزان داده شده است. میانگین نمره های آزمون ۲۵ و واریانس کل ۲۶ محاسبه شده است. سطح دشواری سوال ها نیز یکسان است. ضریب قابلیت اعتماد این آزمون را محاسبه کنید:

$$r_2 = \frac{40}{39} \left(1 - \frac{25(40-25)}{40(26)} \right) = 1.036(1 - 0.36) = 0.66$$

اعتبار(روایی)

اعتبار(روایی) به ارتباط منطقی، بین پرسش های آزمون و مطلب مورد سنجش اشاره دارد. وقتی گفته می شود آزمون، روایی دارد به این معنا است که پرسش های آزمون به طور دقیق آنچه را که مورد نظر می باشد، می سنجد. اعتبار، جنبه های مختلف دارد و ارتباط بین پرسش و آزمودنی با توجه به کلیه جنبه های آن حاصل می شود. در صورتی که این ارتباط وجود نداشته باشد اعتبار به وجود نمی آید.

انواع اعتبار یا روایی ابزار جمع آوری اطلاعات در تحقیقات کمی عبارتند از:

اعتبار محتوایی: آزمون، وقتی اعتبار محتوایی دارد که هدف های آن با محتوای آزمودنی ارتباط داشته باشد. برای مثال، سنجش دقت افراد از طریق آزمون هایی امکان پذیر است که با مفهوم دقت ارتباط داشته باشد.

اعتبار ساخت: اعتبار ساخت، در مورد آزمون هایی لازم است که به منظور سنجش استعداد های کلی و پدیده های دارای ابعاد مختلف طراحی می شوند. برای مثال، هوش پدیده ذهنی است که دارای ابعاد مختلفی از قبیل: سرعت انتقال، عکس العمل متناسب، تشخیص روابط بین پدیده ها و میزان سازگاری می باشد.

اعتبار درونی: این اعتبار با توانا ساختن پژوهشگر در جمع آوری اطلاعات و تجزیه و تحلیل آنها، با حذف کلیه عوامل مداخله گر و تعبیر و تفسیر درست آنها سروکار دارد.

اعتبار بیرونی: این اعتبار به قابلیت تعمیم پذیری یافته های تحقیق ارتباط دارد. به این معنی که آیا نتایج آزمایش را می توان به جامعه ای که نمونه از آن انتخاب شده است، تعمیم داد یا نه. به عبارت دیگر، آیا یافته های پژوهش، معرف و بیانگر شرایط و موقعیت های زمانی و مکانی خاص دیگر هستند؟ گرچه، اعتبار بیرونی با جامعه ای که پژوهشگر قصد دارد نتایج را به آن تعمیم دهد، مربوط است.

اعتبار صوری: اعتبار صوری این مطلب را مد نظر دارد که سؤال‌های آزمون تا چه حد در ظاهر شبیه به موضوعی هستند که برای اندازه‌گیری آن تهیه شده‌اند. در واقع روایی صوری نمی‌تواند نوعی روایی باشد، بلکه تنها یک ویژگی آزمون است که در پاره‌ای مواقع وجود آن مفید است. در بعضی از آزمون‌ها به‌ویژه آزمون‌های استخدامی، اگر آزمون فاقد روایی صوری باشد، آزمون‌شونده ممکن است علاقه‌ای به جواب دادن سؤال‌های آزمون از خود نشان ندهد، زیرا ممکن است چنین تصور شود که آزمون به تصمیم‌های مربوط به استخدام او ربطی ندارد.

اعتبار ملاکی: اعتبار ملاکی به صورت همبستگی نمره‌های آزمون با یک ملاک خارجی که با متغیر مورد سنجش مربوط است، تعریف می‌شود. اعتبار وابسته به ملاک را به اعتبار همزمان و اعتبار پیش‌بین تقسیم می‌کنند. اگر اندازه متغیر ملاک همزمان با اجرای آزمون به دست آید، همبستگی آزمون با متغیر ملاک را اعتبار همزمان می‌نامند. مانند، همبستگی بین نمره‌های هوش و پیشرفت تحصیلی گروهی از دانش‌آموزان در یک زمان معین. در اعتبار پیش‌بین، اندازه متغیر ملاک مدتی پس از اجرای آزمون به دست می‌آید. به عنوان مثال همبستگی بین نمره‌های یک تست هوش با نمره‌های پیشرفت تحصیلی آزمودنی‌ها در یک یا چند سال آینده، اعتبار ملاکی پیش‌بین آزمون هوش نامیده می‌شود.

اعتبار سازه: اعتبار سازه بر این مطلب تأکید می‌کند که آزمون تا چه اندازه سازه نظری یا صفت مورد نظر را اندازه می‌گیرد. این سنجش مستلزم سه اقدام اساسی است: در ابتدا سازنده آزمون باید صفت مورد نظر را به دقت تحلیل کند. در مرحله بعد، چگونگی ارتباط صفت با متغیرهای دیگر را مورد توجه قرار دهد و بعد از طریق آزمایش معلوم کند که آیا این روابط فرضی واقعا وجود دارند یا نه.

اعتبار استقرایی: میزان رابطه شهودی ظاهری اقلام یک آزمون با رفتارهایی که فرض می‌شود، مورد سنجش قرار می‌گیرند.

اعتبار همخوان: اعتبار تثبیت‌شده از طریق مقایسه نتایج آزمونی تازه با نتایج آزمونی که قبلا معتبر شناخته شده است.

اعتبار وفاقی: اعتبار یک آزمون یا اصلی که با نسبت افراد موافق یا معتبر بودن آن تعیین می‌شود.

اعتبار همگرا: میزان همبستگی نمرات یک آزمون با چندین عامل مختلف را گویند. وجود این همبستگی برای اطمینان از این که آزمون آنچه را که باید سنجیده شود می‌سنجد، ضروری است.

اعتبار تعریفی: اعتبار یک آزمون، متکی بر این است که اقلام تشکیل‌دهنده آن طبق تعریف آنچه را که باید سنجیده شوند می‌سنجد.

اعتبار افتراقی: اعتبار یک آزمون، مبتنی بر میزان پیش‌بینی عملکرد افتراقی در دو یا چند تکلیف را گویند.
اعتبار تفکیکی: میزان نقص همبستگی یک آزمون با آزمایشات یا مهارت‌هایی که قرار نیست سنجیده شوند را گویند.

اعتبار تجربی: می‌توان به تجربه نشان داد که یک آزمون آنچه را مورد نظر است می‌سنجد.

اعتبار عاملی: میزان همبستگی نمرات در نتایج دو آزمونی که قرار است چیز واحدی را بسنجد.

اعتبار نمونه‌گیری: می‌زانی که به نظر می‌رسد یک آزمون صفات نمونه‌گیری شده خاص را در زمینه هر آنچه قرار است سنجیده شود، می‌سنجد.

اعتبار وابسته به قوانین طبیعی: می‌زانی که به نظر می‌رسد یک آزمون با توجه به یک نظریه کلی، آنچه را که باید بسنجد، می‌سنجد.

اعتبار ترکیبی: اعتبار یک وسیله آزمایش پیچیده یا مجموعه کاملی از آزمون‌ها، مبتنی بر رابطه بین نمره مرکب که بازنمایی عوامل مختلفی است که در آزمون عملکرد واقعی نمایانده می‌شوند.

اعتبار ویژگی: می‌زانی که یک آزمون هر یک از صفات زمینه‌ای هر آنچه را که قرار است سنجیده شود، می‌سنجد.

اعتبار افزایشی: می‌زانی که یک آزمون از طریق پالایش تدریجی (مثلاً با حذف مواردی که هماهنگ با سایر مواد معتبر نیست) به معیاری قابل اطمینان برای سنجش آنچه مورد نظر است، تبدیل می‌گردد.