

بسم الله الرحمن الرحيم



دانشگاه علوم پزشکی تبریز
واحد توسعه تحقیقات بالینی مرکز آموزشی درمانی کودکان



روایی و پایایی ابزارهای اندازه گیری کمی

دکتر ماهنی رهکار فرشی
استادیار گروه کودکان دانشکده پرستاری مامایی تبریز

روایی یا اعتبار validity

آیا واقعا همان چیزی را اندازه میگیریم
که قصد اندازه گیری آن را داریم؟

(اعتبار) *Validity*

```
graph TD; A["(اعتبار)  
Validity"] --> B["مفهوم اعتبار به این سؤال پاسخ می‌دهد که ابزار اندازه‌گیری تا چه حد خصیصه مورد نظر را می‌سنجد."]; B --> C["نکته"]; C --> D["بدون آگاهی از اعتبار ابزار اندازه‌گیری نمی‌توان به دقت داده‌های حاصل از آن اطمینان داشت."];
```

مفهوم اعتبار به این سؤال پاسخ می‌دهد که ابزار اندازه‌گیری تا چه حد خصیصه مورد نظر را می‌سنجد.

نکته

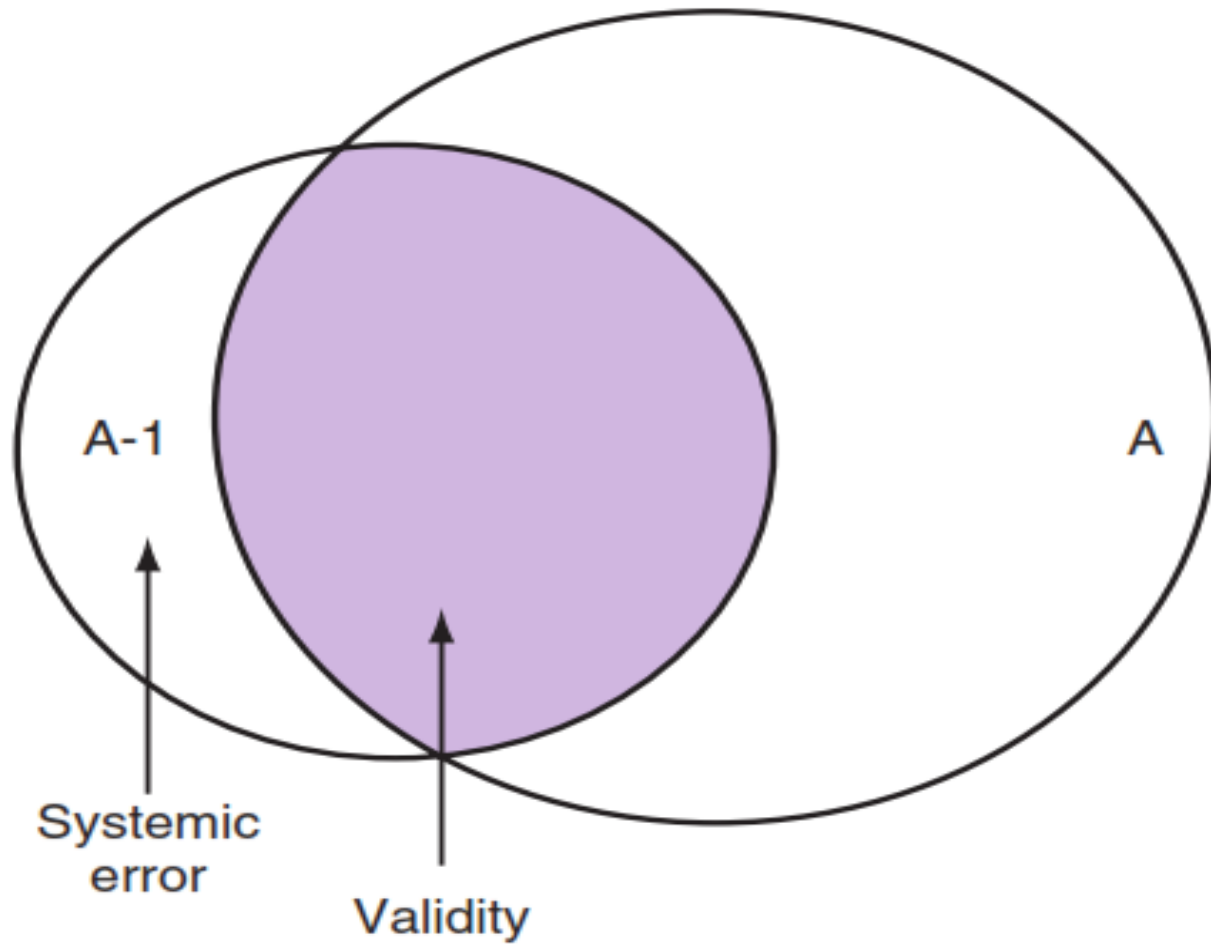
بدون آگاهی از اعتبار ابزار اندازه‌گیری نمی‌توان به دقت داده‌های حاصل از آن اطمینان داشت.

⊙ اعتبار یک پدیده همه یا هیچ نبوده و دارای طیف و درجه است.

⊙ هیچ ابزاری کاملاً معتبر نیست.

⊙ محقق میزان اعتبار ابزار را تعیین می کند نه فقدان یا داشتن آن را

اعتبار ابزار



- معیاری مهم جهت ارزیابی ابزارهای کمی
- ابزار آنچه را که مد نظر است را اندازه بگیرد نه چیز دیگری را.

● مثال: یک آزمون اضطراب روا ابزاری است که تنها اضطراب افراد را اندازه گیری کند نه چیز دیگری را.

انواع روایی ابزارها

۱- اعتبار محتوا Content Validity

۲- اعتبار ملاک و معیار Criterion Validity

۳- اعتبار سازه Construct Validity

۴- اعتبار پیش بین predictive validity

Validity is considered a single broad measurement evaluation that is referred to as **construct validity** and includes various types, such as:

- اعتبار محتوا
- اعتبار از تحلیل عاملی
- همگرا
- اعتبار واگرا
- اعتبار از گروه های متضاد
- اعتبار از پیش بینی رویدادهای آینده
- اعتبار از پیش بینی رویدادهای جاری
- ▣ content validity
- ▣ validity from factor analysis
- ▣ convergent
- ▣ divergent validity
- ▣ validity from contrasting groups
- ▣ validity from prediction of future events
- ▣ validity from prediction current events

سایر اعتبارها شواهدی از
روایی سازه به حساب می
آیند.

روایی صوری Face Validity

• بین سالهای ۱۹۶۰ تا ۱۹۷۰ تنها نوع اعتبار در اکثر مطالعات بود.

• روایی صوری به علت انجام قضاوت ذهنی و نبود دستورالعمل واضح و روشن، به عنوان ضعیفترین نوع روایی تلقی می شود.

• به این مطلب اشاره می کند که ابزار از نظر ظاهری تا چه حد شبیه به موضوعی است که برای اندازه گیری آن تهیه شده است.

توماس (۱۹۹۲) در رابطه با روایی صوری اظهار می دارد:

روایی صوری هر چند که ضامن اندازه گیری دقیق نیست، ممکن است بر انگیزش افراد تحت مطالعه و در نتیجه بر روایی نمرات حاصل از ابزار اندازه گیری تاثیر داشته باشد.

خوانایی ابزار readability

- عنصر ضروری برای روایی و پایایی ابزار
- هرچند بطور رسمی به عنوان مؤلفه اعتبار محتوا شناخته نمی شود، ولی ابزاری که غیر قابل درک است، نمی تواند اعتبار محتوا داشته باشد.
- بیش از ۳۰ فرمول خوانایی وجود دارد که برای ارزیابی مؤلفه های زبانی بکار می رود.

روایی محتوا

• مربوط به این است که روش اندازه گیری تا چه حد کل مؤلفه های مفهوم مورد مطالعه را در بر می گیرد.

• نمونه سؤال های مورد استفاده چقدر نماینده کل سوالاتی هستند که میتوان در این موضوع پرسید؟

• مناسب برای آزمونهای دانش و آگاهی و اندازه گیری خصوصیات روانی-اجتماعی

◎ اعتبار محتوی از سه منبع بدست می آید:

1. the literature
2. representatives of the relevant populations
3. Content experts

۱. مبانی نظری
۲. جامعه آماری مرتبط
۳. دانشمندان و متخصصان

● اعتبار محتوای یک ابزار بر اساس قضاوت ذهنی کارشناسان و متخصصان آن حیطة است که به تحلیل کفایت و پوشش سوالات می پردازند.

● در بررسی روایی محتوا می توان نسبت روایی محتوا CVR (content validity ratio) را برای هریک از گزینه های ابزار بدست آورد.

● مقدار عددی اعتبار محتوی از طریق محاسبه شاخص روایی محتوا CVI (content validity index) بدست می آید.

این موارد با استفاده از توافق کارشناسان بدست می آید.

نسبت روایی محتوا CVR (content validity ratio)

ضرورت			گویه ها
ضروری است essential	مفید است اما ضروری نیست useful	ضرورتی ندارد Not necessary	
			.1
			.2
			.3

$$\text{CVR} = (n_e - N/2)/(N/2)$$

Note: n_e = The number of experts who rated an item as “essential”

N = the total number of experts. *Example:* If 8 of 10 experts rated an item as essential, CVR would be $(8 - 5/5) = 0.60$

CVR Statistic (Content Validity Ratio; Lawshe)

$$\frac{n_e - N/2}{N/2}$$

$$\frac{10 - 5}{5} = 1$$

Excellent item

$$\frac{8 - 5}{5} = .60$$

Possible item for inclusion

$$\frac{5 - 5}{5} = 0$$

Poor item

$$\frac{0 - 5}{5} = -1.0$$

Very poor item

شاخص روایی محتوا CVI (content validity index)

	مرتبط بودن	سادگی	واضح بودن
	۱ مرتبط نیست	۲ نیاز به بازبینی جدی دارد	۳ مبهم است (واضح نیست)
	۴ کاملاً مرتبط است	۵ نیاز به بازبینی جدی دارد	۶ کاملاً واضح است
گویه ها	۷ مرتبط بوده اما نیاز به بازبینی دارد	۸ ساده بوده اما نیاز به بازبینی دارد	۹ واضح بوده اما نیاز به بازبینی دارد
.1.			
.2.			
.3.			

CVI for each scale item is the proportion of experts who rate the item as a 3 or 4 on a 4-point scale. *Example:* If 4 of 6 content experts rated an item as relevant (3 or 4), CVI would be: $4/6 = 0.67$

This item would not meet the 0.83 level of endorsement required to establish content validity using a panel of 6 experts at the 0.05 level of significance. Therefore, it would be dropped

CVI for the entire scale is the proportion of the total number of items deemed content valid. *Example:* If 77 of 80 items were deemed content valid, CVI would be: $77/80 = 0.96$

اعتبار تحلیل عامل

Validity from factor analysis

- رویکردی ارزشمند در تعیین روایی سازه ابزار است.
- تعیین ابعاد مختلف یا زیر مؤلفه های یک پدیده
- گزینه هایی که در ارتباط نزدیک با یکدیگر هستند در یک گروه قرار می گیرند
- محققین می توانند تعداد عامل ها و زیر مؤلفه های هر یک را تعیین کنند.
- گزینه ای که مرتبط با سایر گزینه ها نمی باشد و نمی توان آن را در داخل عاملی گنجانده حذف می گردد.

● تحلیل عامل: روشی برای تعیین دسته سوالات
مربوط به هم در یک مقیاس است.

● تحلیل عامل ممکن است وجود چند عامل را نشان
دهد که بیانگر وجود چند سازه در داخل یک ابزار
است.

● تحلیل عامل يك روش پیشرفته آماری است که از
طریق آن تعداد و ماهیت متغیرهایی را که يك
آزمون اندازه می گیرد مشخص می کنند.

تعیین اعتبار بوسیله تحلیل ساختاری

Validity from Structural Analysis

این روش فراتر از تحلیل عامل به محقق بینش می دهد.
تعیین روابط دقیق هر گزینه با عامل

تحلیل عامل  تعیین می کند کدام گزینه ها در یک گروه قرار می گیرند.

تحلیل ساختاری  تعیین می کند که هر گزینه با گزینه دیگر چه ارتباطی دارد.

همگرایی (تجانس) Convergent Validity

- هنگامی که ابزار جدیدی برای مطالعه ساخته می شود ←
- تعیین اینکه ابزار جدید در سنجش یک سازه تا چه حدی به ابزار قدیمی نزدیک است.
- هر دو مقیاس قدیمی و جدید در مورد یک نمونه به اجرا در می آیند. ← انجام تحلیل همبستگی بین دو ابزار
- اگر بین دو اندازه همبستگی مثبت بالایی وجود داشته باشد، اعتبار هر دو ابزار افزایش می یابد.

واگرایی (عدم تجانس) Divergent Validity

- می توان اعتبار ابزار را با ابزار دیگری که سازه ای متضاد با سازه مورد نظر را اندازه می گیرد، سنجید.
- بهتر است این روش همراه با روش اعتبار همگرایی بطور همزمان انجام شود.
- این رویکرد ترکیبی انجام همزمان روایی convergent و divergent ، multitrait-multimethod (MT-MM) نامیده می شود.
- در روایی واگرا اگر بین دو اندازه، همبستگی منفی بالایی وجود داشته باشد، اعتبار هر دو ابزار افزایش می یابد.

گروه های مقابل یا شناخته شده

Validity from Contrasting (or Known) Groups

از طریق شناسایی گروه هایی که انتظار می رود نمرات بسیار متفاوتی با هم داشته باشند.

اندازه گیری حس تعلق در دانشجویان، بیماران دچار افسردگی شدید و راهبه های بازنشسته

این روش اعتبار ابزار را در مورد نمرات گروه ها متناسب با پیش بینی های اولیه، نشان می دهد.

تعیین اعتبار بوسیله تحلیل افتراقی

Evidence of Validity from discriminant Analysis

- گاهی ابزارهایی برای سنجش سازه های نزدیک به سازه مورد بررسی، مورد استفاده قرار می گیرد.
- اگر ابزار جدید بتواند سازه جدید ولی مرتبط به سازه نزدیک را متمایز کند، اعتبار هر دو ابزار افزایش می یابد.
- برای انجام آزمون تمایز باید هر دو ابزار بطور همزمان در نمونه مشابه اجرا شده و تمایزها مشخص گردد.

پیش بینی وقایع آینده و پیش بینی وقایع مقارن (همزمان)

Validity from Prediction of Future Events and Concurrent Events

❖ توانایی پیش بینی نگرش ها و عملکرد آینده بر مبنای نمرات ابزار بر اعتبار آن می افزاید. صحت اعتبار پیش بین از طریق تحلیل رگرسیون تعیین می شود.

• پیش بینی سطح سلامت افراد در آینده براساس آزمون سطوح استرس این افراد در سه سال گذشته

❖ پیش بینی ارزش فعلی یک ابزار بر مبنای ارزش بدست آمده از مفهوم دیگر

• مقایسه نمره عزت نفس یک نفر بر مبنای نمره آن در ابزار سنجش سازگاری: انتظار می رود کسی که نمره بالایی در ابزار سنجش سازگاری می گیرد، در عزت نفس نیز نمره بالایی بگیرد

درستی نتایج منجر به تأیید اعتبار مقارن هر دو ابزار می شود.

تأیید متوالی اعتبار

successive verification of Validity

پس از طراحی اولیه یک ابزار، محققین دیگر از آن ابزار در مطالعات غیر مرتبط دیگر استفاده می کنند



هر کدام از مطالعات اطلاعاتی را به اعتبار ابزار اضافه می کنند



در طول زمان اعتبار یک ابزار بطور متوالی تأیید می شود.

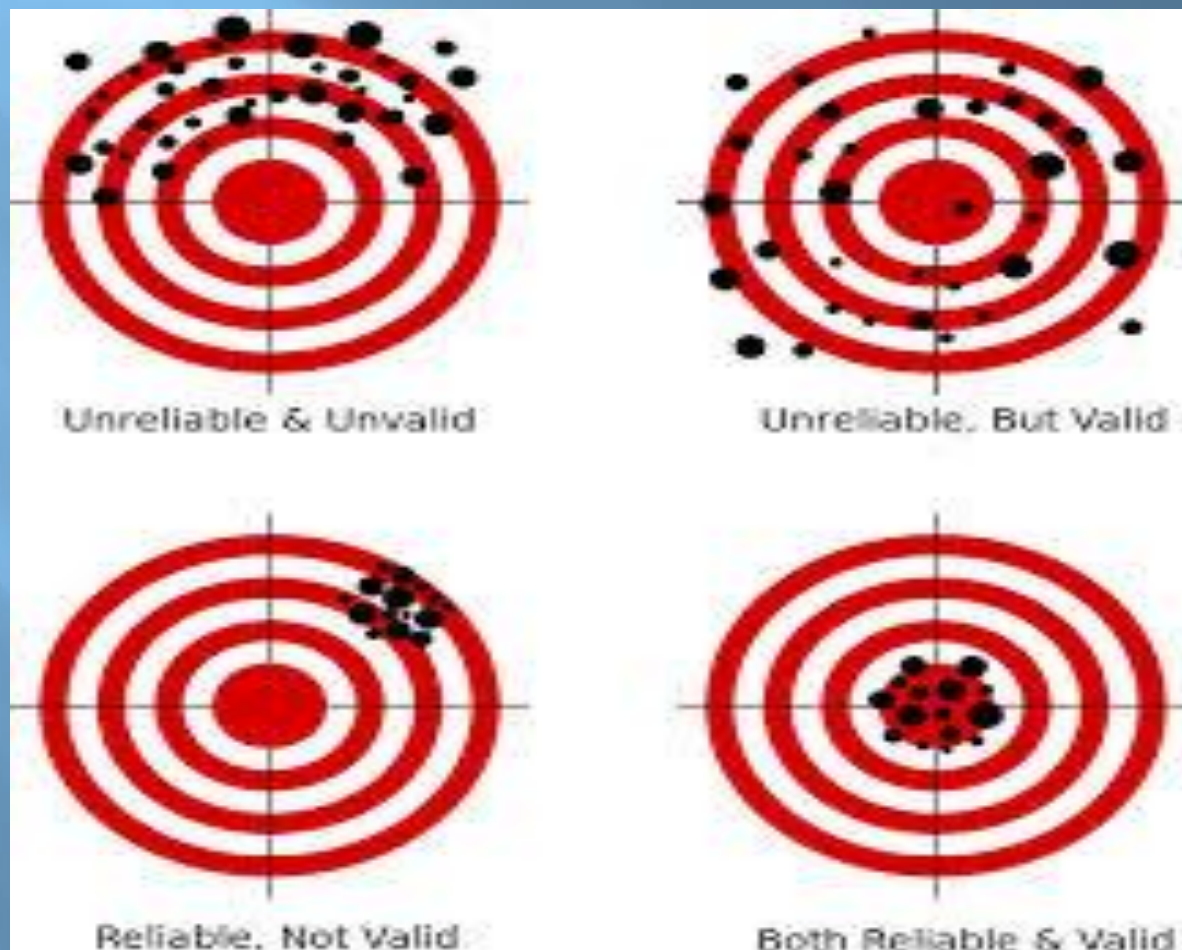
پایایی Reliability

◎ پایایی به دقت، اعتمادپذیری ثبات، یا تکرار پذیری نتایج آزمون اشاره می کند.

◎ پایایی یک ابزار کمی معیار مهمی برای ارزیابی دقت و کیفیت آن است و ثباتی است که با آن یک ابزار، صفت را اندازه می گیرد.

◎ هر چه تغییرپذیری ابزار در سنجش های تکراری یک صفت کمتر، پایایی آن بیشتر

روایی بدون پایایی حاصل نمی شود



- ❖ If the instrument has low reliability values, it cannot be valid because its measurement is inconsistent and has high measurement error (DeVon et al., 2007; Waltz et al., 2010).
- ❖ An instrument that is reliable cannot be assumed to be valid for a particular study or population. You need to determine the validity of the instrument you are using for your study, which you can accomplish in a variety of ways.

انواع پایایی

۱- ثبات stability

۲- هم ارزی Equivalence reliability

- روش پایایی اندازه گیران interrater reliability
- روش فرم های موازی یا هم ارز

alternate-forms reliability or parallel-forms reliability

۳- روشهای همسانی درونی homogeneity/Internal consistency

- روش دو نیمه کردن آزمون
- روش ضریب آلفای کراباخ
- روش کودر- ریچاردسون

ثبات

○ آزمون مجدد test-retest

○ فرض آزمون: عامل مورد اندازه گیری بین دو نوبت اندازه گیری ثابت می ماند و تغییر نمی کند و هر تغییری ناشی از خطای تصادفی است.

○ کاربرد ابزار روی یک نمونه افراد در دو زمان جداگانه

- variability of the variable
- complexity of the measurement Process
- characteristics of the participants

○ فاصله دو هفته تا یک ماه برای ابزارهای قلم و کاغذی

○ برای اندازه های فیزیکی و تکنولوژیکی می توان آزمون و آزمون مجدد را پشت سر هم انجام داد.

○ تحلیل همبستگی بین نمرات دو تست (coefficient of stability)

○ r ضریب پایایی: بین صفر تا یک (هر چه عدد بالاتر ثبات ابزار اندازه گیری بیشتر)

○ مناسب برای ویژگی ها و خصوصیات نسبتاً پایدار که در طول زمان تغییر نکند. (مثل شخصیت، هوش)

هم ارزی روش پایایی اندازه گیران

✓ مناسب برای ابزارهای مشاهده ای دارای ساختار

✓ پایایی مشاهده گران:

- دو یا چند نفر بطور مستقل، واقعه مشابه را بر اساس اصول مشاهده و ثبت می کنند.
- یک واقعه در دو زمان توسط یک مشاهده گر مطالعه و ثبت می شود.

همسانی بین نمرات دو مشاهده گر مستقل به یک پدیده

Number of agreements ÷ number of possible agreements

- اگر بیش از دو مشاهده گر بکار گرفته شود: استفاده از روش آماری ضریب آلفا
- برای آزمون تفاوت بین مشاهده گران: استفاده از آنالیز واریانس
- ارزش پایایی مشاهده گران بهتر است از ۰/۹ یا ۹۰ درصد بالاتر باشد.

هم ارزی

روش فرم های موازی یا هم ارز

- ✓ برای مقایسه دو ابزار قلم کاغذی
- ✓ دو آزمون موازی یا معادل برای یک محتوی یا موضوع تهیه می کنند و آنها را در فاصله زمانی کوتاهی به یک گروه واحد از آزمون شوندگان می دهند.
- ✓ متداول نیست
- ✓ نشان دادن اینکه یک محقق در دو آزمون یک محتوی را می سنجد.

روشهای همسانی درونی یا پیوستگی داخلی

◎ تاکید بر یکنواختی ماده ها یا اجزای تشکیل دهنده یک ابزار دارد.

◎ ابزار با پایایی ثبات داخلی یعنی همه زیر مجموعه ها صفت مشابهی را اندازه بگیرند.

بیانگر همبستگی گزینه های یک ابزار است.

◎ مناسب برای آزمون های قلم کاغذی و مقیاس هایی با سوالات کوتاه و چند گزینه ای

روش دو نیمه کردن آزمون

- یکی از قدیمی ترین روشهای بررسی ثبات درونی است.
 - آزمون مورد نظر را یک بار با گروه واحدی از آزمون شوندگان اجرا می کنیم و پس از اجرا، آن را به دو نیمه تقسیم می نماییم .
در این روش بهترین راه دونیمه کردن آزمون این است که همه سؤال های فرد را یک آزمون حساب آوریم و همه سؤال های زوج را آزمون دیگری بدانیم .
 - نمره دهی به دو گروه سوالات و محاسبه ضریب پایایی و مقایسه نمرات دو نیمه تست .
- استفاده از فرمول همبستگی اسپیرمن براون
- ضریب پایایی بالا یعنی دو نیمه تست واقعاً صفت مشابهی را اندازه میگیرند.

روشهای همسانی درونی

روش ضریب آلفای کراباخ

- تخمینی از همبستگی تمام نیمه های ابزار
- تمام آیتم های ابزار یک مفهوم را می سنجند
- رویه آماری برای بررسی ثبات درونی داده های فاصله ای یا نسبی
- مناسب برای وقتی که یک آزمون از تعدادی خرده آزمون یا بخش های مختلف تشکیل شده است.

Cronbach's alpha coefficients can range from 0.00, indicating no internal consistency or reliability, to 1.00, indicating perfect internal reliability with no measurement error. Alpha coefficients of 1.00 are not obtained in study results because all instruments have some measurement error.

(0.80), (0.70 to 0.79), (0.60 to 0.69), (0.90 to 0.99)

این روش در شرایطی بکار می رود که: نمرات سؤال ، نه به صورت دو ارزشی صفر و یک، بلکه به صورت چند ارزشی (مثلا کاملا موافق تا کاملا مخالف از ۱ تا ۵) تعیین می شود.

روشهای همسانی درونی

روش کودر-ریچاردسون

فرمول کودر-ریچاردسون برای آزمون هایی است که، برای پاسخ آن ها می توان دو حالت در نظر گرفت.

If the data are dichotomous, such as a symptom list that has responses of present or absent, the Kuder-Richardson formulas (KR 20 or KR 21) can be used to calculate the internal consistency of the instrument

Other approaches to testing internal consistency are:

- (1) Cohen's kappa statistic, which determines the percentage of agreement with the probability of chance being taken out.
- (2) correlating each item with the total score for the instrument.
- (3) correlating each item with each other item in the instrument.

This procedure, often used in instrument development, allows researchers to identify items that are not highly correlated and delete them from the instrument.

A misty autumn forest scene. A large, gnarled tree stands on the left, its branches reaching out. The ground is covered in fallen orange and yellow leaves. A fallen log lies in the foreground. The background is shrouded in mist, with more trees visible in the distance. The overall mood is serene and quiet.

با تشکر