

# آشنایی با طراحی و روانسنجی ابزار در مطالعات علوم پزشکی

مهر ماه ۱۴۰۳

دانشگاه علوم پزشکی اصفهان / گروه آمار مدیریت آمار و فناوری اطلاعات

# عناوین مطالب

- آشنایی با مفاهیم ابزار و روانسنجی ابزار
- آشنایی با مراحل ساخت ابزار
- آشنایی با مراحل روانسنجی ابزار

# مفاهيم "ابزار" و "روانسنجی ابزار"

# مفهوم ابزار پژوهش و ابزار اندازه گیری

- ابزاری که برای جمع‌آوری، اندازه‌گیری و تجزیه و تحلیل داده‌های مربوط به موضوع پژوهش مورد استفاده قرار می‌گیرد، ابزار پژوهش نام دارد.
- ابزارهای پژوهش می‌توانند انواع مصاحبه‌ها، آزمون‌ها، پرسشنامه‌ها یا حتی چک لیست‌ها باشند.
- انتخاب ابزار مناسب بستگی به هدف و نوع مطالعه دارد:
  - نوع مطالعه: کیفی / کمی / ترکیبی
  - هدف مطالعه: استخراج مولفه‌ها / آگاهی‌سنجی / نگرش‌سنجی / ارزشیابی و ...

# مفهوم روانسنجی ابزار

- بررسی کیفیت ابزار یا به عبارتی کیفیت داده های جمع آوری شده توسط ابزار
- ویژگی های روان سنجی به اعتبار و پایایی ابزار اندازه گیری اشاره دارد.
- بسته به نوع ابزار استفاده شده، مراحل روانسنجی ممکن است تغییر کند.

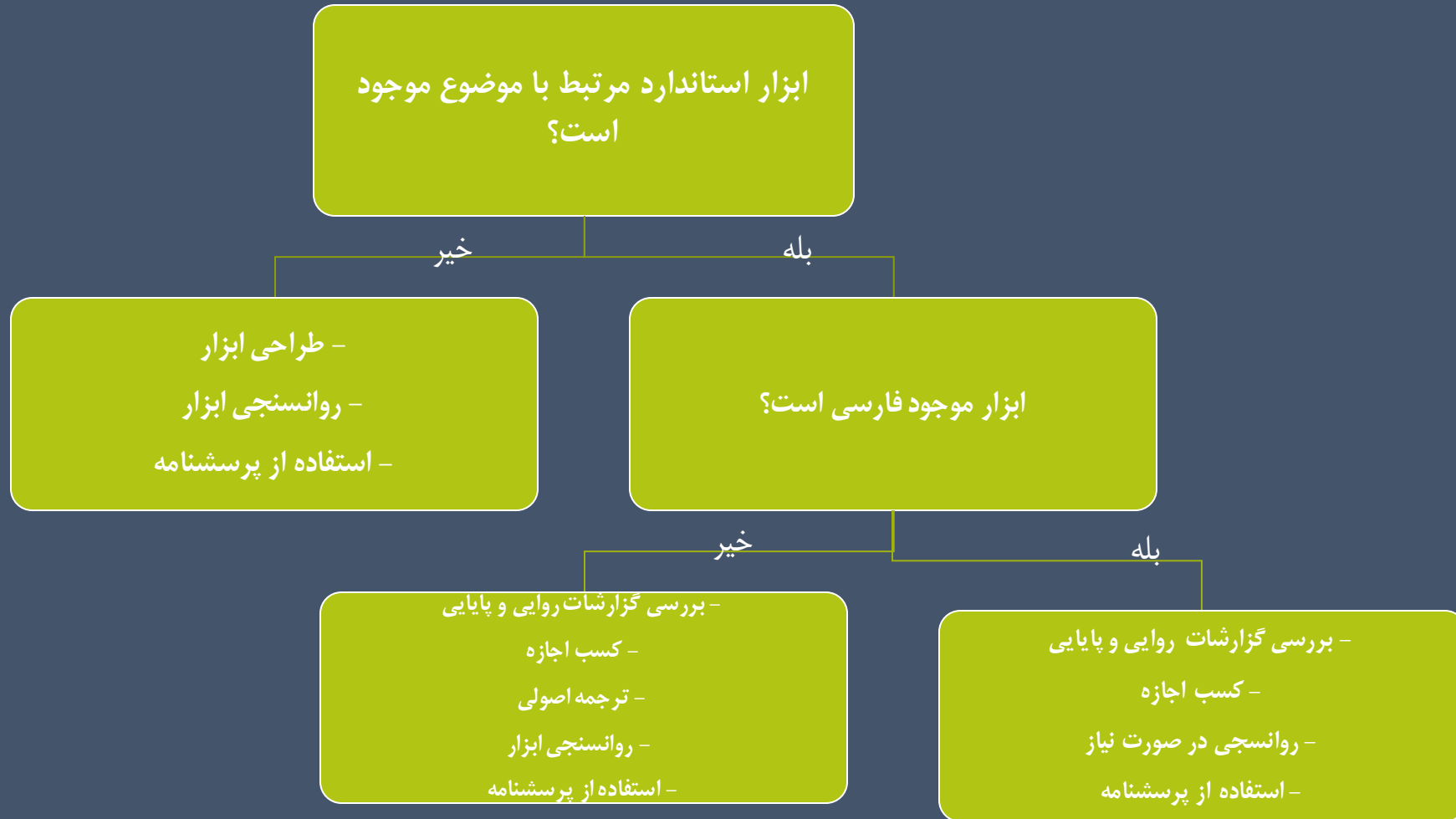
## • رویکردهای روانسنجی ابزار

- نظریه کلاسیک آزمون (CTT)
- نظریه سوال پاسخ (IRT)
- ترکیبی

## • انواع ابزار از این دیدگاه

- ابزارهای استاندارد
- ابزارهای محقق ساخته

# به طور خلاصه...



# مراحل ساخت ابزار

پرسشنامه

# مراحل ساخت پرسشنامه

(۱) تعیین هدف پرسشنامه!

(۲) تدوین ابزار شامل تولید آیتم های ابزار (گویه‌ها/سوالها) ابزار و تعیین مقیاس آنها

- استفاده از رویکردهای قیاسی، استقرایی و یا ترکیبی از هر دو جهت تعیین حوزه ی محتوایی (شامل همه ی چیزهایی که به هدف پرسشنامه مربوط میشود)، طبقات یا سازه ها یا مولفه های اصلی، عوامل اصلی تاثیر گذار

- مرور متون (مقاله ها، پایان نامه ها، کتابها و پژوهش ها...)

- مصاحبه با متخصصین حوزه و با جامعه ی هدف

- بحث گروهی متمرکز / بارش افکار / اجرای روش دلفی

- و نهایتاً تولید آیتم ها



# مراحل ساخت پرسشنامه / تدوین ابزار / انتخاب مقیاس پاسخ ها

- مقیاس گویند ها
  - سوالات باز پاسخ (وقت گیرند، برای مخاطب معمولاً دشوار است، تجزیه و تحلیل نیز دشوارتر است)
  - سوالات بسته پاسخ (دو گزینه ای، چند گزینه ای، مقیاس های درجه بندی و...)
- نوع سوالها به هدف پرسشنامه و اینکه چه خصیصه ای را قرار است اندازه گیری کنیم بستگی دارد.
- سوالهای دو گزینه ای (صحیح - غلط) یا چند گزینه ای بیشتر برای مواردی نظیر اندازه گیری پیشرفت تحصیلی یا آگاهی سنجی و ... مناسب هستند.
- مقیاسهای درجه بندی برای سنجش نگرش، حالات خلقی، انگیزش، سنجش شخصیت و ... مناسبند.
- گزینه ها به صورتی باشد که پاسخ دهنده دست کم بتواند یکی از گزینه ها را انتخاب کند.

# مراحل ساخت پرسشنامه / تدوین ابزار / انتخاب مقیاس پاسخ ها / طیف لیکرت



- مقیاس یا طیف لیکرت نوعی مقیاس رتبه بندی است که به طور گسترده در انواع تحقیقات استفاده می شود . هدف اصلی استفاده از این مقیاس تعیین نظرات، نگرش ها و رفتارهای افراد مختلف است.

- دو قالب متداول از پاسخ ها در مقیاس لیکرت وجود دارد: پنج گزینه ای و هفت گزینه ای

# مراحل ساخت پرسشنامه / تدوین ابزار / انتخاب مقیاس پاسخ ها / طیف لیکرت

• سوالات لیکرت برای سنجش متغیرهای مختلفی مانند فراوانی، کیفیت، موافقت و غیره کاربرد دارند. طیف لیکرت اندازه‌گیری مفاهیم غیر عددی و کیفی را ممکن می‌کند.

- فراوانی – بررسی رفتار، عادت‌ها، علائم بیماری
  - خیلی زیاد / زیاد / معمولی / کم / خیلی کم
- کیفیت – سنجش کیفیت و انتظار از کیفیت
  - خیلی ضعیف / ضعیف / معمولی / خوب / خیلی خوب
- موافقت / مخالفت
  - کاملاً موافقم / موافقم / نظری ندارم / مخالفم / کاملاً مخالفم
- رضایت – نظرسنجی‌های بیماران / ارزیابی دوره‌های آموزشی
  - کاملاً راضی / راضی / نظری ندارم / ناراضی / کاملاً ناراضی
- اهمیت
  - بسیار با اهمیت / با اهمیت / نظری ندارم / بی اهمیت / بسیار بی اهمیت
- شدت یا میزان ویژگی – وجود یا عدم وجود صفت

# مراحل ساخت پرسشنامه / تدوین ابزار / انتخاب مقیاس پاسخ ها / طیف لیکرت

## • مزایا

- کمی کردن نتایج و امکان انجام آنالیزهای آماری

- سهولت برای پاسخ دهندگان

- گزینه ای مناسب در طراحی سوالات بسته پاسخ

- امکان استفاده در انواع پرسشنامه ها

## • معایب

- معانی متفاوت گزینه ها برای افراد (گاهی، به ندرت و ...)

- گرایش به مرکز

- محدودیت نسبت به باز پاسخ ها

- سختی طراحی سوال

# ادامه مراحل ساخت پرسشنامه

- انجام اصلاحات سوالها جهت ورود به مراحل روانسنجی

- (۳) روانسنجی ابزار


- انجام مراحل و تایید روایی ابزار

- انجام مراحل و تایید پایایی ابزار

- (۴) تبیین قاعده نمره گذاری پرسشنامه

- (۵) تنظیم پرسشنامه

# مراحل روانسنجی ابزار



# مفهوم روایی

روایی یک ابزار (validity) میزان صحت نتایج حاصل از آن را نشان میدهد به عبارتی نشان میدهد ابزار تا چه اندازه چیزی را که قصد اندازه گیری آن را داشته است به درستی سنجیده است؟

آیا آن چیزی را که شما فکر می کنید اندازه می گیرد اندازه گیری می کند؟

آیا یافته ها منعکس کننده نظرات، نگرش ها و رفتارهای جامعه هدف هستند؟



# مفهوم پایایی

پایایی ابزار (Reliability) نشان می‌دهد ابزار اندازه‌گیری در شرایط یکسان تا چه اندازه نتایج یکسانی به دست می‌دهد.

ثبات در طول زمان

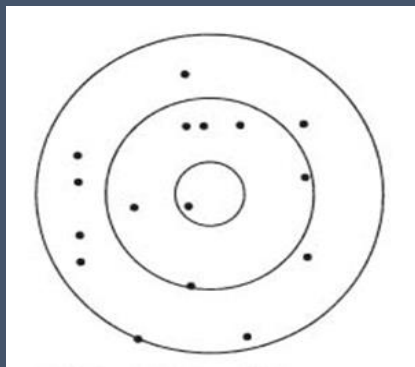
ثبات در امتیازها

ثبات در آیتم‌ها

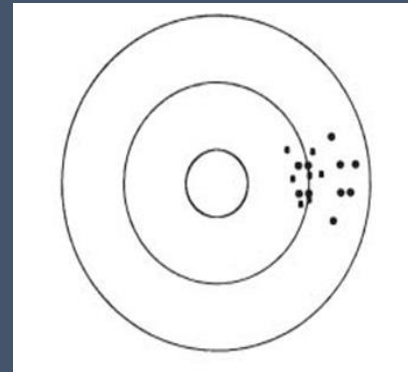


# روایی و پایایی ابزار

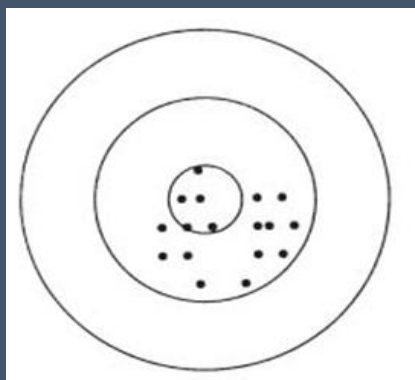
روایی و پایایی ندارد



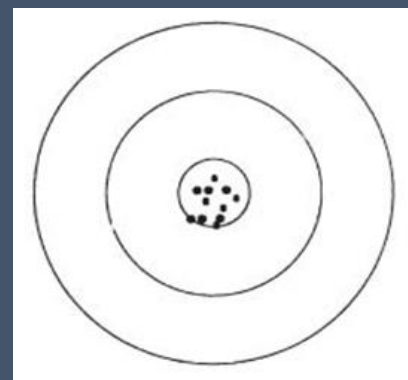
روایی ندارد ولی پایایی دارد



تا حدودی روایی دارد ولی چندان پایایی نیست



روایی و پایایی دارد



# روایی و پایایی ابزار

- پایایی جزئی از روایی است.
- بررسی پایایی نسبتاً ساده است، در حالیکه بررسی روایی بسیار دشوارتر و اغلب مملو از تئوری است.
- لزوماً لازم نیست هر بار که از یک روش اندازه گیری استفاده می شود، روایی و پایایی آن را نشان دهید.
- گاهی اوقات ابزارهایی وجود دارد که محققان قبلی روایی و پایایی آنها را نشان داده‌اند و شما می‌توانید آنها را به سادگی استفاده کنید.



**آشنایی با انواع روایی و روش بررسی آنها**

# عناوین مطالب

- روایی ترجمه
- روایی مرتبط با محتوا
- روایی تصویری
- روایی محتوایی
- روایی مرتبط با ملاک
- روایی پیش بین
- روایی همزمان
- روایی مرتبط با سازه
- روایی همگرا
- روایی واگرا یا تشخیصی
- روایی عاملی یا روایی ساختاری



# روایی ترجمه

Translation validity

# روایی ترجمه

- ترجمه فرایندی است که طی آن پیامی از زبان مرجع به زبان هدف قابل انتقال میشود. این فرایند نیازمند رعایت قواعد و اصول خاص خود است.
- در کنار ترجمه پرسشنامه، تطبیق فرهنگی معانی عبارات بسیار حائز اهمیت است؛ زیرا برای آن که بتوان پرسشنامه را با بستر فرهنگی هر جامعه تطبیق داد، باید برابری های معنایی و محتوایی عبارات انجام شود.
- ارزش های منعکس شده به وسیله ابزار و مفاهیم ممکن است از فرهنگی به فرهنگ دیگر بسیار متفاوت باشد؛ بنابراین کیفیت ترجمه و روایی نسخه ترجمه شده ابزار نقش مهمی در معتبر سازی نسخه ترجمه شده ایفا مینماید.

# مراحل روایی ترجمه

- رویکردهای مختلفی برای ترجمه ابزارها اتخاذ میشود اما تمامی آنها از اصول مشترکی برخوردارند.

- **مراحل ترجمه (وایلد و همکاران)**

(۱) ترجمه پرسشنامه از زبان اصلی به زبان هدف (forward translation)

(۲) ترکیب و تلفیق ترجمه های اولیه به یک ترجمه واحد

(۳) برگرداندن نسخه نهایی ترجمه شده از زبان هدف به زبان اصلی (backward translation)

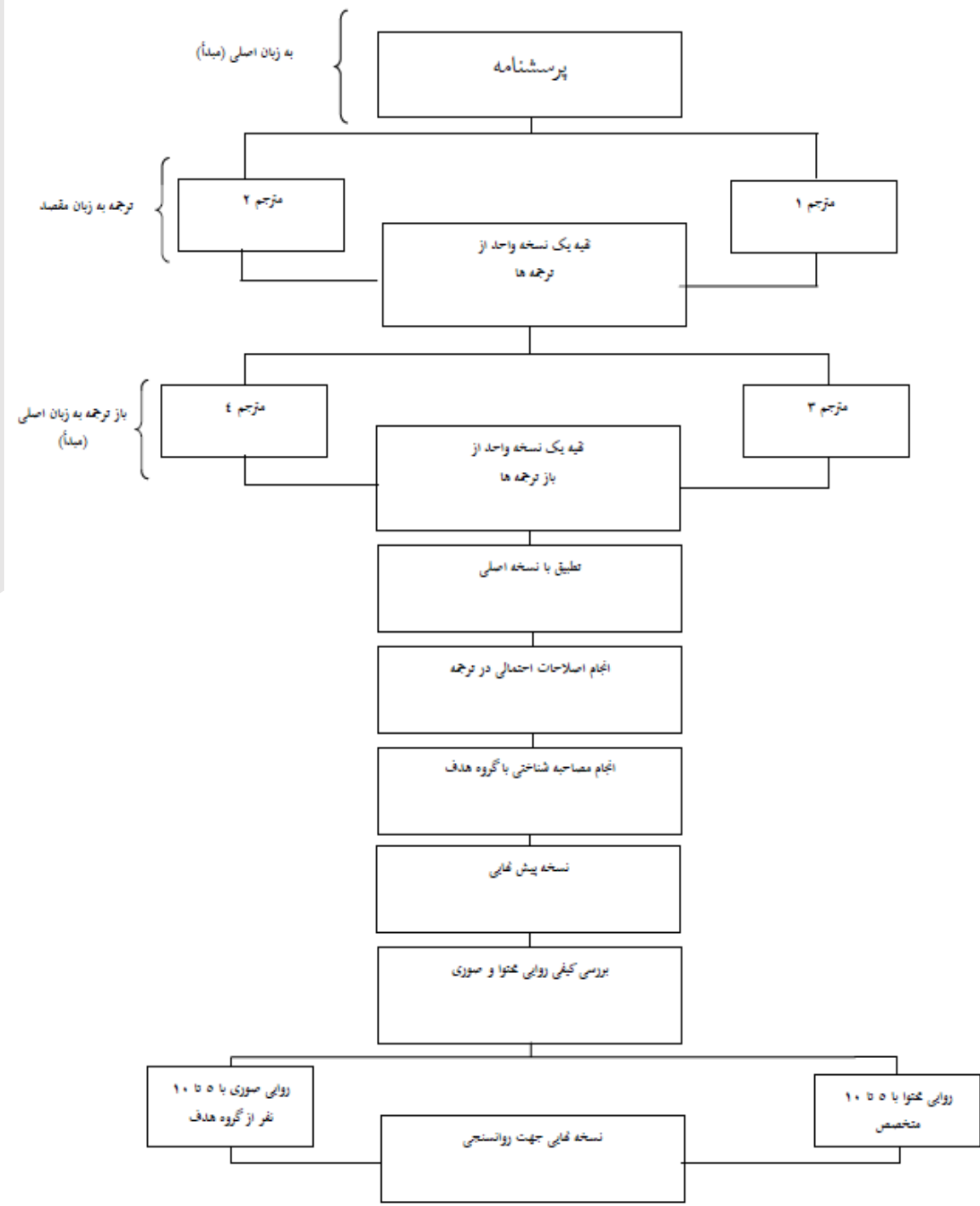
(۴) بازنگری نسخه ترجمه شده از زبان هدف به زبان اصلی

(۵) انجام مطالعه مقدماتی

(۶) اصلاح و جمع بندی

(۷) تعیین روایی و پایایی پرسشنامه یا ابزار ترجمه شده

(۸) ارائه ی گزارش نهایی.





# روایی مربوط به محتوا

Content-Related Validity

# روایی مربوط به محتوا

- هدف از روایی محتوا کسب اطمینان از توانایی ابزار برای سنجش پدیده یا مفهومی است که ادعای سنجش آن را دارد.

- آیا گویه های پرسشنامه ویژگی مورد نظر را به طور جامع پوشش می دهند؟

- مشخص کردن و توجه به هدف ابزار به طور دقیق و مداوم

- مشخص کردن و توجه به گروه هدف به طور دقیق

- توجه به سادگی و قابل فهم بودن آیتم ها

- انواع روایی مربوط به محتوا

- روایی محتوایی

- روایی صوری

# روایی محتوایی

- روایی محتوایی با استفاده از نظر متخصصان در زمینه موضوع و به دو صورت کیفی و کمی انجام می شود.
- هر چه افراد انتخاب شده در این گام در مفهوم مورد سنجش ابزار دارای دانش و تجربه بیشتری بوده و به موضوع اشراف بیشتری داشته باشند فرایند تأمین روایی محتوا مطمئن تر پیش می رود.
- **روایی محتوا به روش کیفی:**
  - درخواست مطالعه دقیق ابزار از متخصصان منتخب
  - ارائه دیدگاه های اصلاحی خود را در مورد رعایت دستور زبان، استفاده از کلمات مناسب، قرارگیری گویه ها در جای مناسب
  - اظهار نظر در رابطه با مرتبط بودن گویه ها با هدف، واضح بودن، ساده و شفاف بودن گویه ها
  - انجام اصلاحات براساس نظرات ارائه شده

# روایی محتوایی

## روایی محتوا به روش کمی:

- برای این منظور از دو ضریب نسبت روایی محتوا (CVR) و شاخص روایی محتوا (CVI) استفاده می‌شود.

(۱) CVR در رابطه با ضرورت وجود گویه است. برای تعیین CVR از متخصصان درخواست میشود تا هر گویه را بر اساس طیف سه قسمتی زیر بررسی نمایند:

- ضروری است (نمره ۳)
- مفید است، ولی ضرورتی ندارد (نمره ۲)
- ضرورتی ندارد (نمره ۱)

# روایی محتوایی / کمی

- سپس پاسخها مطابق فرمول زیر محاسبه گردید:

$$CVR = \frac{n_E - \frac{N}{2}}{\frac{N}{2}}$$

- در این رابطه،  $n_E$  تعداد متخصصینی است که به گزینه‌ی ضروری پاسخ داده‌اند و  $N$  تعداد کل متخصصین است. اگر مقدار محاسبه شده از مقدار جدول لاوشه بزرگ‌تر بود، اعتبار محتوای آن گویه پذیرفته می‌شود.

# جدول لاوشه

تعداد متخصصان	کمترین مقدار CVR قابل قبول
۵	۰/۹۹
۶	۰/۹۹
۷	۰/۹۹
۸	۰/۷۵
۹	۰/۷۸
۱۰	۰/۶۲
۱۱	۰/۵۹
۱۲	۰/۵۶
۱۳	۰/۵۴
۱۴	۰/۵۱
۱۵	۰/۴۹
۲۰	۰/۴۲
۲۵	۰/۳۷
۳۰	۰/۳۳
۳۵	۰/۳۱
۴۰	۰/۲۹

# روایی محتوایی / کمی

(۲) CVI در رابطه با میزان مربوط بودن، ساده بودن و وضوح گویه-ها است. برای تعیین CVI از متخصصان درخواست میشود تا هر گویه را بر اساس طیف چهار قسمتی زیر بررسی نمایند:

- مرتبط بودن یا اختصاصی بودن:

- کاملاً مرتبط (۴)، مرتبط (۳)، نسبتاً مرتبط (۲) و مرتبط نیست (۱).

- ساده بودن:

- کاملاً ساده (۴)، ساده (۳)، نسبتاً ساده (۲) و ساده نیست (۱).

- شفافیت و وضوح:

- کاملاً واضح (۴)، واضح (۳)، نسبتاً واضح (۲) و واضح نیست (۱).

# روایی محتوایی / کمی

- امتیاز شاخص روایی محتوا برای هر گویه (I-CVI) به وسیله تقسیم تعداد متخصصین موافق با عبارات دارای رتبه‌ی ۳ و ۴ بر تعداد کل متخصصین محاسبه می‌شود.

نتیجه	مقدار CVI
گویه مناسب است	۰/۷۹ و بالاتر
نیاز به اصلاح و بازنگری	بین ۰/۷۰ تا ۰/۷۹
قابل قبول نیست و حذف گویه	۰/۶۹ و پایین تر



# روایی صوری / کیفی

- روایی صوری، عبارت است از میزان مناسب بودن ظاهر ابزار برای جمع آوری اطلاعات مورد نظر به ویژه از دیدگاه پاسخگویان.
- **روایی صوری کیفی:** مصاحبه چهره به چهره با چند نفر از گروه هدف
  - بیان درک خود از گویه توسط گروه هدف
  - سوال درباره سطح دشواری (دشواری درک گویه‌ها و کلمات)، میزان تناسب (تناسب و ارتباط مطلوب گویه‌ها با پرسشنامه) و ابهام (احتمال وجود برداشت‌های اشتباه از گویه‌ها و یا وجود نارسایی در معانی کلمات)
- این مصاحبه توسط محقق اصلی که اشراف کامل بر گویه‌ها دارد انجام شود.
- گویه‌ها با توجه به توصیه‌های این گروه ویرایش می‌گردد. و این روند تا جایی پیش می‌رود که درک گویه‌ها برای افراد ساده شده و تغییر جدیدی پیشنهاد نشود.

# روایی صوری / کمی

- روایی صوری به روش کمی: سوال از تعدادی افراد گروه هدف در رابطه با اهمیت هر گویه براساس طیف زیر:
- کاملاً مهم است (۵)، تا اندازه‌ای مهم است (۴)، در حد متوسط مهم است (۳)، اندکی مهم است (۲)، مهم نیست (۱).
- در این روش نسبت شرکت‌کنندگانی که گویه را با اهمیت می‌دانند (فراوانی ۴ و ۵ به درصد) در متوسط اهمیت گویه (میانگین نمره کسب شده) ضرب می‌شود.  
$$\text{میانگین اهمیت} \times (\%) \text{ فراوانی} = \text{نمره تأثیر گویه}$$
- در صورتی که امتیاز تاثیر از ۱/۵ بیشتر شود، آیتم برای تحلیل‌های بعدی مناسب تشخیص داده شده و حفظ خواهد شد.

# روایی وابسته به ملاک

- این روایی نشان می‌دهد که ابزار ساخته شده تا چه حد با یک استاندارد بیرونی مرتبط است که معمولاً با محاسبه ضریب همبستگی بین آنها سنجیده میشود.
- نیازمند به وجود یک معیار استاندارد بیرونی در حال حاضر یا در آینده هست.
- اگر همبستگی بالایی دارند، به این معنی است که آزمون (پیش‌بینی‌کننده) معتبر است! - نشان می‌دهد که اندازه‌گیری ما می‌تواند چیزی را به درستی پیش‌بینی یا برآورد کند که ما از لحاظ نظری معتقدیم که باید بتواند آن را پیش‌بینی کند.

## • انواع روایی ملاکی:

- پیش‌بین
- همزمان

# روایی ملاکی / پیشین

- 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.

مثال:

- پیشبینی قبولی و یا رد شدن دانشجویان در پایان ترم براساس نمره میان ترم
- آزمون های انتخاب کارکنان در برابر معیارهای عملکرد شغلی
- آزمون داخلی انتخاب تیم المپیاد مدرسه یا دانشگاه
- سنجش این نوع روایی زمان بر و هزینه بر است.

# روایی ملاکی / همزمان

- هنگامی برآورد میشود که معیار (استاندارد) در حال حاضر وجود دارد. مثال:
- بررسی همبستگی پاسخ کارمندان در رابطه با میزان حقوق دریافتی با معیار مدارک حسابداری
- آزمون مهارت‌های مربوط به «سخنرانی عمومی» دانش‌آموزان در مقابل عملکرد واقعی (سخنرانی عمومی) دانش‌آموزان در داخل یا خارج از کلاس
- همبستگی امتحان آیین نامه و امتحان عملی رانندگی
- تهیه ی نسخه خلاصه از پرسشنامه های استاندارد

# روایی ملاکی / همزمان

- بر خلاف اعتبار پیش‌بینی‌کننده، که در آن اندازه‌گیری دوم دیرتر اتفاق می‌افتد، اعتبار همزمان به اندازه‌گیری دوم تقریباً در همان زمان نیاز دارد.
- در مطالعاتی که در مورد این نوع اعتبار انجام می‌شود، دو آزمون تقریباً همزمان انجام می‌شود و محققان ارزیابی می‌کنند که آیا آزمون مورد مطالعه می‌تواند به عنوان جایگزین بالینی برای آزمون **Gold Standard** استفاده شود یا خیر.
- این رویکرد در شرایطی مفید است که یک ابزار جدید یا آزمایش نشده به طور بالقوه کارآمدتر، اداره کردن آن آسان‌تر، عملی‌تر، یا ایمن‌تر از روش دیگر شناخته شده‌تر است و به عنوان یک جایگزین پیشنهاد می‌شود.
- استفاده از ضرایب همبستگی و مقادیر حساسیت و ویژگی

# روایی مربوط به سازه

## Construct-Related Validity

۹/۲۵/۲۰۲۴

۳۹

# روایی سازه

- ساختن: کنار هم قرار دادن مواد یا قطعات، به طور منظم، به منظور ساخت یک بنای مشخص (ساختمان، پل، و غیره...)
- سازه در ابزارسازی یک مفهوم نظری و انتزاعی است که ابزار اندازه گیری برای توصیف آن ساخته میشود. مثل: انگیزه، عصبانیت، شخصیت، هوش، کیفیت زندگی، سبک زندگی و ...
- بررسی کیفیت سازه فیزیکی در مقابل سازه انتزاعی؟
- روایی مرتبط با سازه در روان سنجی به این اشاره دارد که آیا ابزار ساختار مفهومی تئوری شده (مثلاً هوش) را اندازه گیری می کند یا با آن ارتباط دارد .



# روایی سازه

## • انواع روایی سازه:

- روایی همگرا (Convergent Validity)
- روایی تشخیصی یا واگرا (Discriminant Validity)
- روایی ساختاری یا تحلیل عاملی (Factor Analysis)
- تحلیل عاملی اکتشافی (Exploratory Factor Analysis)
- تحلیل عاملی تاییدی (Confirmatory Factor Analysis)

# روایی سازه / همگرا

- همبستگی هر گویه با گویه های هم حیطه
- همبستگی هر سوال با نمره کل پرسشنامه
- همبستگی نمره سازه با ابزارهای مشابه

وقتی گروه هدف، ابزار طراحی شده و ابزاری که سازه مشابه را می سنجد تکمیل کنند، بین نمرات این دو باید همبستگی وجود داشته باشد. وقتی ابزار جدید با ابزار موجود، که برای اندازه گیری همان سازه ساخته شده است، دارای همبستگی باشد، آزمون جدید نیز برای اندازه گیری آن سازه رواست.

**Table 2.** Correlation matrix for different domains of WHOQOL-100 and WHOQOL-BREF (n=60)

Domain		WHOQOL-100					
		Physical	Psychological	Level of Ind.	Social relation.	Environment	Spirituality /religion.
<b>WHOQOL-BREF</b>							
Physical health	r <sup>*</sup>	0.759	0.268	0.694	0.465	0.358	0.456
	P <sup>**</sup>	<0.001	0.037	<0.001	<0.001	0.005	<0.001
Psychological health	r	0.354	0.739	0.603	0.589	0.550	0.671
	P	0.006	<0.001	<0.001	<0.001	<0.001	<0.001
Social relationships	r	0.216	0.195	0.427	0.665	0.541	0.484
	P	0.098	0.132	0.001	<0.001	<0.001	<0.001
Environmental health	r	0.081	0.017	0.297	0.297	0.683	0.413
	P	0.537	0.896	0.020	0.020	<0.001	0.001

\* Pearson's correlation coefficient

\*\* Two-tailed P-value

# روایی سازه / تشخیصی یا واگرا

- همبستگی ضعیف بین سؤالات مربوط به یک حیطة با دیگر حیطة ها (ضریب همبستگی یک سؤال با حیطة های دیگر باید کمتر از ضریب همبستگی آن سؤال با حیطة های مربوط به خودش باشد).
- عدم همبستگی با ابزارهایی که پدیده های نامربوط را می سنجند.
- اختلاف معنادار نمره ی حاصل از ابزار بین گروه های متفاوت
- انواع آزمونهای آماری از جمله  $t$ -test، ANOVA و رگرسیون در این زمینه قابل استفاده هستند.

**Table 6.** The comparison of different facets and domains of WHOQOL-100 between healthy and disabled participants

		Healthy		Disabled		P*	P**
		Mean	SD	Mean	SD		
<b>Facet</b>	1 Pain and discomfort	13.18	3.15	11.75	2.77	<0.001	<0.001
	2 Energy and fatigue	13.71	2.15	13.47	2.02	0.069	0.065
	3 Sleep and rest	13.71	3.34	12.27	3.27	<0.001	<0.001
	4 Positive affect	11.03	2.45	12.02	2.21	<0.001	<0.001
	5 Thinking, Learning, memory and concentration	13.22	2.61	13.07	2.57	0.360	0.337
	6 Self-esteem	12.00	2.40	13.21	2.32	<0.001	<0.001
	7 Body image and appearance	13.13	2.39	12.60	2.37	<0.001	0.004
	8 Negative affect	11.42	2.36	11.37	2.24	0.731	0.913
	9 Mobility	13.88	3.08	12.86	2.34	<0.001	<0.001
	10 Activities of daily living	11.81	2.06	11.74	2.11	0.596	0.799
	11 Dependence on medication or treatments	15.94	2.92	14.67	3.50	<0.001	<0.001
	12 Working capacity	14.62	3.11	13.44	3.66	<0.001	<0.001
	13 Personal relationships	13.73	2.57	13.79	2.69	0.718	0.563
	14 Social support	11.55	3.40	12.31	3.59	0.001	0.002
	15 Sexual activity	12.57	2.49	11.94	3.07	<0.001	<0.001
	16 Physical safety and security	13.05	2.94	12.82	2.19	0.161	0.996
	17 Home environment	12.67	2.36	12.46	3.45	0.261	0.438
	18 Financial resources	12.58	2.23	11.98	2.74	<0.001	0.681
	19 Health and social care: accessibility and quality	11.73	2.78	12.36	2.97	0.001	<0.001
	20 Opportunities for acquiring new information ...	11.74	2.81	10.59	2.67	<0.001	0.003
	21 Participation in and opportunities ...	10.14	2.80	12.34	2.16	<0.001	<0.001
	22 Physical environment	12.07	2.33	12.89	3.50	<0.001	<0.001
	23 Transportation	13.12	3.08	14.65	2.88	<0.001	<0.001
	24 Spirituality/religion/personal beliefs	14.65	3.39	14.66	3.50	0.888	0.915
	25 Overall Quality of life and general health	13.98	2.89	13.63	3.19	0.069	0.112
<b>Domain</b>	<b>Physical</b>	13.53	2.25	12.50	2.15	<0.001	<0.001
	<b>Psychological</b>	12.46	1.08	12.25	1.02	<0.001	<0.001
	<b>Level of independence</b>	14.06	1.76	13.18	2.29	<0.001	<0.001
	<b>Social relationship</b>	12.98	1.83	12.68	1.19	<0.001	<0.001
	<b>Environment</b>	12.70	1.76	12.30	1.86	<0.001	<0.001
	<b>Spirituality/religion/personal beliefs</b>	14.65	3.39	14.66	3.50	0.888	0.915

# روایی سازه / تحلیل عاملی یا ساختاری

- تحلیل عاملی به معنی تعیین ابعاد پرسشنامه هست که آن را روایی ساختاری نیز می گویند.

- اگر هیچ پیش فرضی در مورد ابعاد تشکیل دهنده ابزار نداشته باشیم تحلیل عامل اکتشافی انجام می شود ولی اگر پیش فرض پذیرفته شده قبلی در مورد ابعاد داشته باشیم باید تحلیل عامل تأییدی انجام داد. (عامل / مولفه / حیطه / متغیر)

طراحی ابزار (اکتشافی / تأییدی)

بومی سازی ابزار (تأییدی / اکتشافی)

- قبل از انجام تحلیل عاملی اجرای پایلوتی جهت سنجش پایایی ابزار توصیه میشود. (در بحث پایایی بیشتر توضیح داده خواهد شد)

# عواملی که باعث کاهش روایی میشوند

- اهداف نامشخص
- آیتم های آزمایشی ضعیف ساخته شده اند
- مشکل در خواندن واژگان و ساختار جمله
- اقلام تست نامناسب برای نتایج مورد سنجش
- ابهام در اظهارات
- تست هایی که خیلی کوتاه هستند
- محدودیت های زمانی نامناسب
- چیدمان نامناسب اقلام
- سطح دشواری نامناسب
- الگوهای قابل شناسایی پاسخ







# پایایی (Reliability)

# عناوین مطالب

- تعریف نظری پایایی / برآورد ضریب پایایی
- انواع پایایی در نظریه آزمون کلاسیک

ثبات در طول زمان (Stability Over Time)

- تست مجدد (Test-Retest) / فرم های موازی (Parallel Forms)

ثبات امتیازات (Stability of Scores)


- ارزیاب درونی (Intra-Rater) / بین ارزیاب (Inter-Rater)

ثبات یا هماهنگی آیتم ها (Stability Over Items)

- تقسیم نصف (Split Half) / آلفای کرونباخ (Cronbach's Alpha) / کرودر - ریچاردسون (Kruider-Richardson)

# مرور مفاهیم روایی و پایایی

- روایی یک ابزار (validity) میزان صحت نتایج حاصل از آن را نشان میدهد به عبارتی نشان میدهد ابزار تا چه اندازه چیزی را که قصد اندازه گیری آن را داشته است به درستی سنجیده است؟
- پایایی ابزار (Reliability) نشان می دهد ابزار اندازه گیری در شرایط یکسان تا چه اندازه نتایج یکسانی به دست می دهد.
- قابلیت اطمینان به معنای تکرارپذیری یا ثبات یا سازگاری است.
- معیاری قابل اعتماد در نظر گرفته می شود که بارها و بارها نتیجه یکسانی را به ما بدهد (با فرض اینکه چیزی که اندازه گیری می کنیم تغییر نمی کند!)



# انواع پایایی در نظریه آزمون کلاسیک

# ثبات در طول زمان (Stability Over Time)

تست مجدد (Test-Retest)

فرم های موازی (Parallel Forms)

# تست مجدد (Test-Retest)

- برای ارزیابی سازگاری یک معیار از زمانی به زمان دیگر استفاده می شود.
- به عنوان مثال گروهی از پاسخ‌دهندگان برای نمرات IQ مورد آزمایش قرار می‌گیرند: هر پاسخ‌دهنده دو بار مورد آزمایش قرار می‌گیرد و این دو آزمون با یک ماه فاصله انجام میشوند.
- این رویکرد فرض می‌کند که هیچ تغییر اساسی در ساختار اندازه‌گیری شده بین دو موقعیت وجود ندارد.
- هر چه فاصله زمانی کوتاهتر باشد، همبستگی بیشتر است. هر چه فاصله زمانی بیشتر باشد، همبستگی کمتر است

# تست مجدد / همبستگی درون طبقاتی (ICC)

- ICC تفاوت ضریب همبستگی پیرسون و

x	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱
y	۱۰	۸	۹	۷	۶	۴	۵	۳	۲	۱

- $r = ۰.۹۵۲$  ,  $ICC = ۰.۹۷۵$

X	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱
y	۱۰۰	۸۰	۹۰	۷۰	۶۰	۴۰	۵۰	۳۰	۲۰	۱۰

- $r = ۰.۹۵۲$  ,  $ICC = ۰.۸۱۷$

# ویژگی های تست مجدد

- اثر انتقالی : افرادی که پاسخ های دفعه قبل را به خاطر می آورن
- قابلیت اطمینان می تواند بیش از حد تخمین بزند (پاسخ های گذشته را تکرار کنید) / یا دست کم گرفتن (پاسخ های گذشته را تصحیح کنید)
- اثر تمرین : تکرار آزمایش باعث بهبود عملکرد می شود
- به عنوان مثال، نمرات IQ ممکن است در آزمون بعدی بالاتر باشد.
- تغییر در موضوعات . مقدار زمان مجاز بین اقدامات بسیار مهم است.
- تغییرات بزرگ / آموزش / تبلیغات و ...
- ریزش نمونه : افرادی که برای آزمایش مجدد حضور ندارند.
- اعتراض : افرادی که به آزمون مجدد اعتراض دارند ممکن است پاسخ های اشتباهی داشته باشند



# فرم های موازی

- به دنبال قابلیت اطمینان از طریق معادل سازی بین دو نسخه از یک آزمون و مقایسه نتایج دو آزمون
- اجرای دو شکل مختلف از یک آزمون با استفاده از آیتم های متفاوت برای اندازه گیری ویژگی های یکسان (مشابه در محتوا، سطح دشواری، ترتیب، و غیره)
- بهتر از تست مجدد است زیرا می توان آن را در همان روز انجام داد.
- ساختن آزمون های معادل مشکل است (آزمون ها باید دارای تعداد آیتم های یکسان، روش امتیازدهی یکسان، محتوای یکسان و سطح دشواری آیتم باشند).

مثال: امتحان های مقاطع یکسان در یک مدرسه

# ثبات آیتم ها یا سازگاری داخلی (Stability Over Items)

دو نیمه کردن (Split Half)

کرودر- ریچاردسون (Kruider-Richardson)

آلفای کرونباخ (Cronbach's Alpha)

# ثبات آیتم ها یا سازگاری داخلی

- سازگاری درونی معمولاً معیاری است که بر اساس همبستگی بین آیتم‌های مختلف در یک آزمون (یا همان خرده مقیاس در یک آزمون بزرگتر) محاسبه میشود.
- سازگاری درونی اندازه‌گیری می‌کند که آیا چندین آیتم که برای اندازه‌گیری یک ساختار کلی تهیه شده‌اند، نمرات مشابهی دارند یا خیر.
- به عنوان مثال ، اگر پاسخ دهنده با جملات "من دوست دارم دوچرخه سواری کنم" و "من در گذشته از دوچرخه سواری لذت می‌بردم" موافق است و با عبارت "من از دوچرخه متنفرم" مخالفت کرد ، این می‌تواند نشان دهنده سازگاری داخلی خوب آزمون باشد.
- هدف در طراحی یک ابزار قابل اعتماد این است که نمرات موارد مشابه مرتبط باشد (از لحاظ درونی سازگار باشد)، اما هر کدام اطلاعات منحصر به فردی را نیز ارائه دهند.

# ارزیابی سازگاری داخلی

- بسته به ابزار موجود روش های مختلفی برای ارزیابی سازگاری داخلی وجود دارد:
- روش دو نیمه کردن
- شاخص کودر – ریچاردسون
- شاخص آلفای کرونباخ

# ارزیابی سازگاری داخلی / روش دو نیمه کردن

- ابزار را در نمونه مورد مطالعه استفاده کنید.
- سوالات را به دو قسمت تقسیم کنید و آنها را جداگانه نمره دهید.
- اگر واریانس این دو قسمت برابر بود از فرمول اسپیرمن و براون استفاده میشود:

$$R_{SB} = \frac{2r_{half}}{1 + r_{half}}$$

- اگر واریانس این دو قسمت برابر نبود، از فرمول کرونباخ و فلنگان استفاده میشود:

$$R_{CF} = \frac{2[s_T^2 - (s_1^2 + s_2^2)]}{s_T^2}$$

# ارزیابی ثبات درونی / شاخص کودر- ریچاردسون-۲۰

- فرمول ۲۰-Kuder -Richardson (KR-۲۰) برای اولین بار در سال ۱۹۳۷ منتشر شد، معیاری برای سازگاری درونی برای سازهایی با انتخاب های دو حالتی است.

$$KR20 = \frac{K}{K - 1} \left[ 1 - \frac{\sum_{i=1}^K p_i q_i}{s_T^2} \right]$$

- K تعداد اقسام است،
- $p_i$  نسبت پاسخ درست به سوال i است ،
- $q_i$  نسبت پاسخ نادرست به سوال i است .

# ارزیابی ثبات درونی / شاخص کودر – ریچاردسون – ۲۱

- فرمول (KR-۲۱) Kuder-Richardson (۲۱) زمانی استفاده می شود که بتوانیم فرض کنیم که همه موارد تقریباً دارای سختی مساوی هستند.

$$KR21 = \frac{K}{K-1} \left[ 1 - \frac{M(K-M)}{KS} \right]$$

- K تعداد اقسام است،
- M میانگین نمرات خام از کل آزمون است،
- S واریانس نمرات خام از کل آزمون است.

# ارزیابی همسانی درونی / شاخص آلفای کرونباخ

- شاخص آلفای کرونباخ اولین بار توسط لی کرونباخ در سال ۱۹۵۱ آلفا نامگذاری شد. می توان آن را با فرمول محاسبه کرد:

$$\alpha = \frac{K}{K - 1} \left[ 1 - \frac{\sum_{i=1}^k s_i^2}{s_T^2} \right]$$



# ارزیابی همسانی درونی / شاخص آلفای کرونباخ

- آلفای کرونباخ به طور کلی با افزایش همبستگی بین آیتم های آزمون افزایش می یابد.
- تابعی از تعداد آیتم هاست و ممکن است دچار بیش برآورد یا کم برآورد شود.
- یک کران پایین محسوب میشود.
- از آنجایی که همبستگی بین آیتم های آزمایشی زمانی به حداکثر می رسد که همه آیتم ها ساختار یکسانی را اندازه گیری می کنند، اعتقاد بر این است که آلفای کرونباخ به طور غیرمستقیم درجه ای را نشان می دهد که مجموعه ای از آیتم ها یک سازه پنهان تک بعدی را اندازه گیری می کنند.
- به عنوان یک قاعده کلی، قبل از استفاده از یک ابزار، به پایایی  $0.7$  یا بالاتر (که در یک نمونه آزمایشی به دست می آید) نیاز داریم.

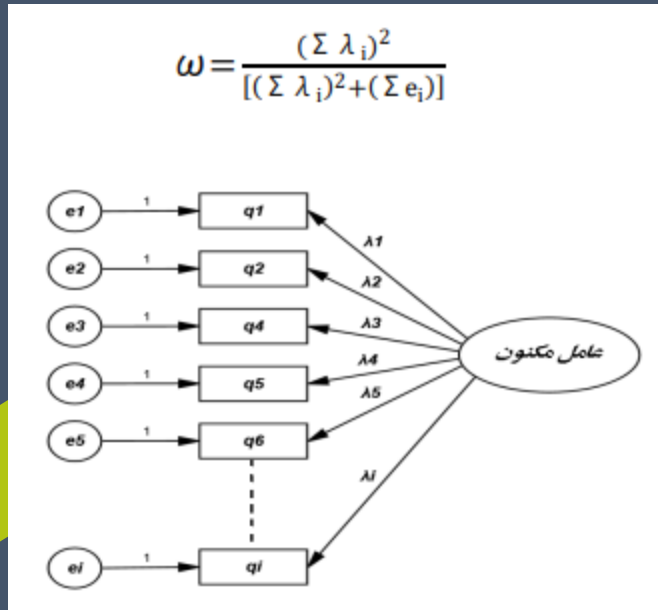
# راه های افزایش ثبات درونی

- دو راه برای افزایش سازگاری داخلی یک ابزار وجود دارد:

- افزایش تعداد گویه ها

- حذف موارد نامناسب

- ضریب امگا مک دونالد جایگزین ضریب آلفا کرونباخ



# ثبات امتیازها (Stability of Scores)

ارزیابی درونی (Intra-Rater)

بین ارزیاب (Inter-Rater)

# بین ارزیاب (Inter-Rater)

- قابلیت اطمینان بین ارزیاب ، توافق بین ارزیاب یا تطابق میزان توافق بین ارزیاب ها است.
- فرض بر این است که ارزیاب های معتبر با یکدیگر موافق هستند. یا ارزیاب های قابل اعتماد در مورد اینکه کدام عملکرد بهتر و کدام بدتر است توافق دارند.
- در پالایش ابزارهایی که به قضاوت انسانی نیاز دارند مفید است مانند چک لیست ها. اگر ارزیاب های مختلف موافق نباشند، یا ابزار پایا نیست یا ارزیاب ها نیاز به آموزش مجدد دارند.

# ارزیابی بین ارزیاب

- تعدادی آمار وجود دارد که می توان از آنها برای تعیین پایایی بین ارزیاب ها استفاده کرد.
- آمارهای مختلف برای انواع مختلف اندازه گیری مناسب است. مثلا:
  - همبستگی بین ارزیاب،
  - ضریب همبستگی
  - همبستگی درون رده ای
  - احتمال توافق مشترک
  - کاپا کوهن
  - کاپا فلیس

# سپاس از همراهی شما

