



خلاصه روش تحقیق



دکتر نورالدین نخستین انصاری
استاد گروه فیزیوتراپی
دانشکده توانبخشی
دانشگاه علوم پزشکی تهران



فراموش نکنید:

- داده های متریک و داده های نامتریک
- داده های متریک برای بررسی مقادیر و اندازه استفاده می شوند.
- داده های نامتریک قابل سنجش نیستند و برای توصیف و طبقه بندی استفاده می شوند.

مقیاس اندازه گیری

چهار مقیاس اندازه گیری اصلی

Nominal

اسمی ■

Ordinal

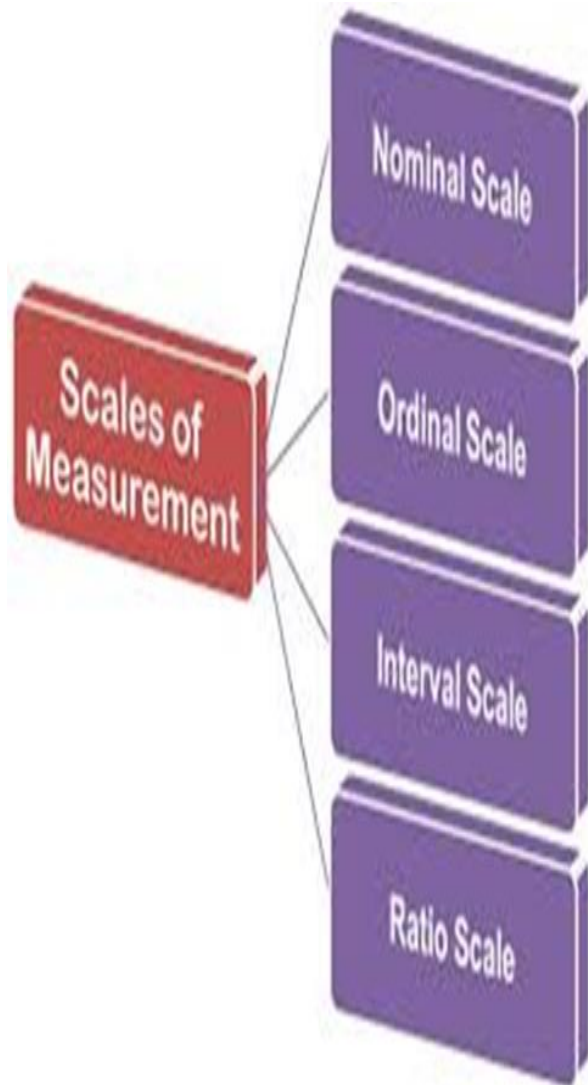
رتبه ای ■

Interval

فاصله ای ■

Ratio

نسبتی ■



Nominal Scales

مقیاس های اسمی

✓ فقط برای طبقه بندی کیفی استفاده می شوند

✓ فاقد نقطه صفر مطلق

✓ نمی توان با سکانس کمی مرتب کرد

✓ امکان انجام اعمال ریاضی با آنها نیست



Nominal Scales



مقیاس های اسمی

- جنسیت
- مذهب
- گرایش سیاسی
- محل تولد
- وضعیت تاهل
- رنگ مو / چشم
- وضعیت استخدام

مقیاس های رتبه ای

Ordinal Scales

- سنجش یک متغیر از لحاظ اسمی و رتبه هر دو
- نمایانگر رتبه متغیرها
- روشن نبودن تفاوت بین طبقات یا رتبه ها، **چه مقدار؟**
- اطلاعات در باره موقعیت نسبی، اما نه فاصله بین رتبه ها یا طبقات

مقیاس های رتبه ای

Ordinal Scales

- جایگاه دونده ها یا شنا گران در یک مسابقه
- نفر اول عملکرد بهتری نسبت به بقیه داشت، اما نمی دانیم چقدر
- زمان طی شده، نیاز به سطح یا مقیاس اندازه گیری دیگر



مقیاس های متریک

Metric Scales

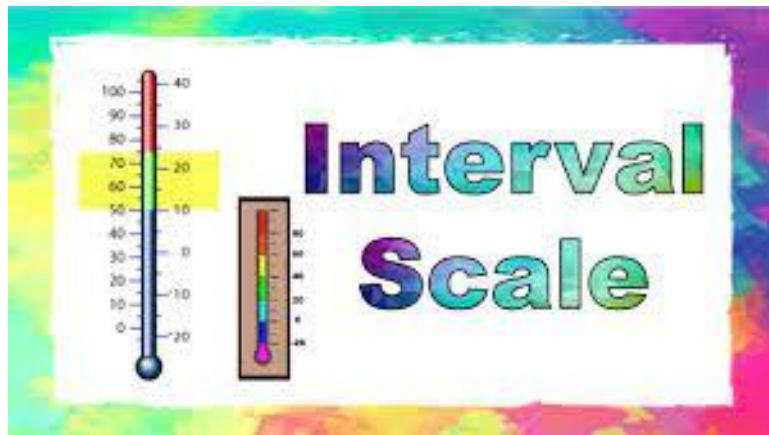


- ماهیت کمی دارند
- مقیاس فاصله ای
- مقیاس نسبتی

مقیاس های فاصله ای

Interval Scales

- اطلاعاتی در باره رتبه و فاصله
- فاصله مساوی بین اعداد
- نبودن نقطه صفر مطلق (قراردادی است)

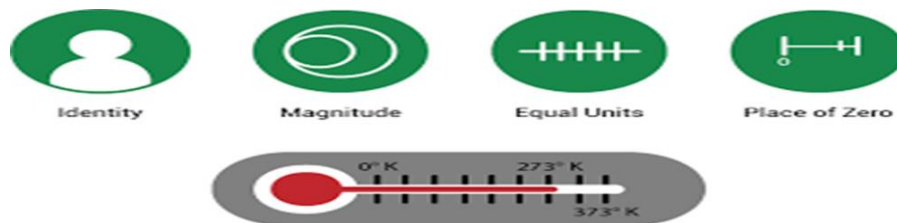


- انجام جمع و تفریق
- مقیاس فارنهایت یا سلسیوس

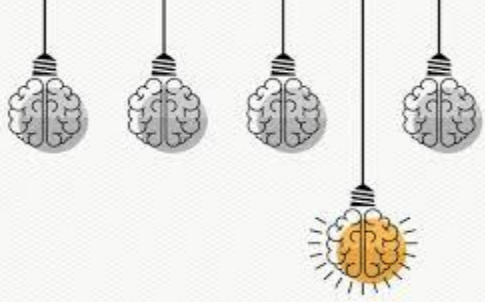
مقیاس های نسبتی

Ratio Scales

- بالاترین سطح اندازه گیری
- ویژگی های آن مشابه مقیاس فاصله ای است، بجز آنکه نقطه صفر مطلق نیز دارد.
- انجام همه اعمال ریاضی و تکنیک های آماری پیچیده



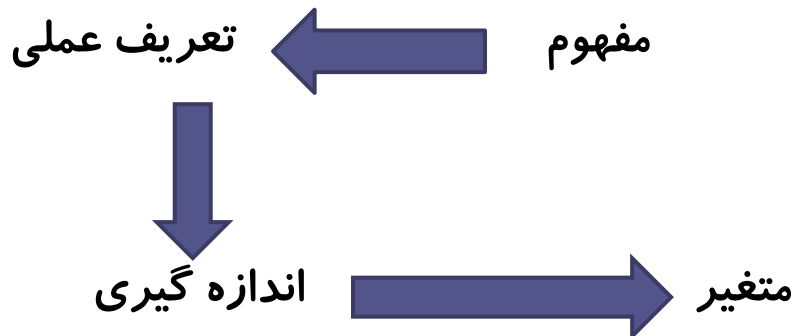
■ قد، وزن، و زمان



جنبه های روان سنجی

ضرورت تعریف عملی در پژوهش برای کمک به سنجش مفاهیم انتزاعی

عملی کردن از طریق اندازه گیری





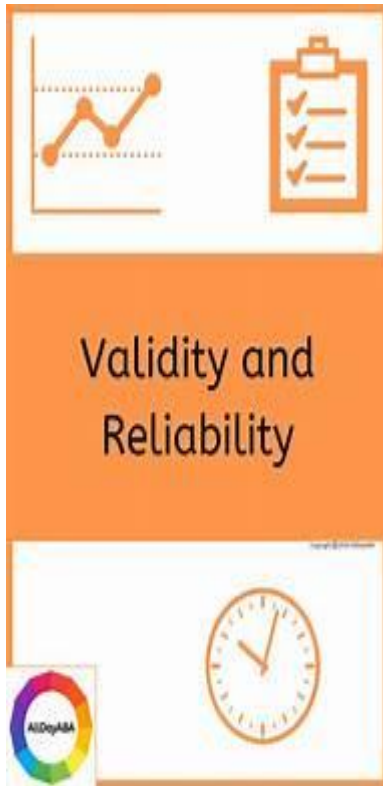
درمان جدید برای افسردگی

- تعریف عملی افسردگی
- چگونگی اندازه گیری یا سنجش آن
- ملاحظه همه راه های بالقوه برای عملی کردن و اندازه گیری آن
- ملاحظه بهبودی، تغییر نکردن، بدتر شدن بیماران پس از دریافت مداخله جدید

چگونگی سنجش بهبودی؟

بسته به تعریف عملی

- بهبودی با مشاهده کاهش نمره در ارزیابی افسردگی
- کاهش علائم در مصاحبه
- مشاهده افزایش سطح فعالیت
- یا مشاهده دو یا همه این شاخص ها
- انتخاب با پژوهشگر، سوال پژوهش، در دسترسی منابع، در دسترسی تکنیک ها و استراتژی های اندازه گیری



پایایی

اعتبار

رایج ترین و مهم ترین مفاهیم مربوط به

ابزار ارزیابی و سایر استراتژی های اندازه گیری

پایایی و اعتبار

و رابطه آنها با اندازه گیری

- پایایی، پیوستگی یا ثبات نمره های بدست آمده از یک معیار یا تکنیک اندازه گیری در طی زمان و در همه محیط ها و شرایط
- پایایی اندازه گیری، شانس کمتر برای کسب نمره ناشی از فاکتورهای تصادفی و خطای اندازه گیری

نمرات بدست آمده از جمع آوری داده در معرض خطای اندازه گیری هستند.

نمره بدست آمده

نمره واقعی True score

خطای اندازه گیری Measurement error

- همه معیارها خطا دارند
- هر قدر روش یا ابزار پایا تر باشد، احتمال اثر عوامل مختلف بر دقت اندازه گیری ها کمتر
- ✓ ارزیابی شخصیت با یک ابزار عینی و استاندارد دو بار، فاصله یک هفته



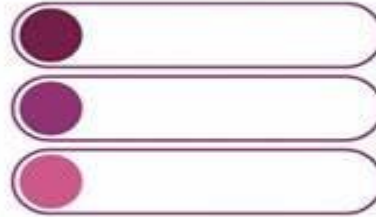
تغییر قابل ملاحظه نتایج
دلیل؟
ناپایایی احتمالی ابزار تست

■ هنگام انتخاب یک ابزار یا رویکرد اندازه گیری،
پایایی ضروری و مهم، اما کافی نیست.

■ اهمیت اعتبار به عنوان بخشی از استراتژی کلی اندازه گیری



Validity



اعتبار validity

تست یا استراتژی اندازه گیری

- آیا ابزار یا رویکرد اندازه گیری آنچه را که قرار است اندازه گیری کند، اندازه گیری می کند؟
- ابزار اندازه گیری معتبر است اگر نمایانگر سازه مورد نظر بوده و آن را دقیق ارزیابی می کند

- اعتبار و پایایی مفاهیم مرتبط با یکدیگر هستند.
- اندازه گیری نمی تواند معتبر باشد مگر این که پایا باشد.

پایایی نشان اعتبار است





فراموش نکنید:

اعتبار

■ میزان دقت تست و ابزار اندازه گیری است

اعتبار به این سوال پاسخ می دهد:

«آیا ابزار یا رویکرد اندازه گیری آن چیزی که فرض است اندازه گیری کند، اندازه گیری می کند؟»

مهم

✓ اندازه گیری دقیق سازه ها و متغیرها جزء مهم در یک پژوهش

اعتبار رویکردهای اندازه گیری جنبه مهم پروسه کلی دیزاین



تعیین اعتبار استراتژی اندازه گیری: چگونه؟

■ مثل پایایی، بر اساس رابطه کمی یا کیفی بین تست یا استراتژی اندازه گیری و معیار مستقل بیرونی

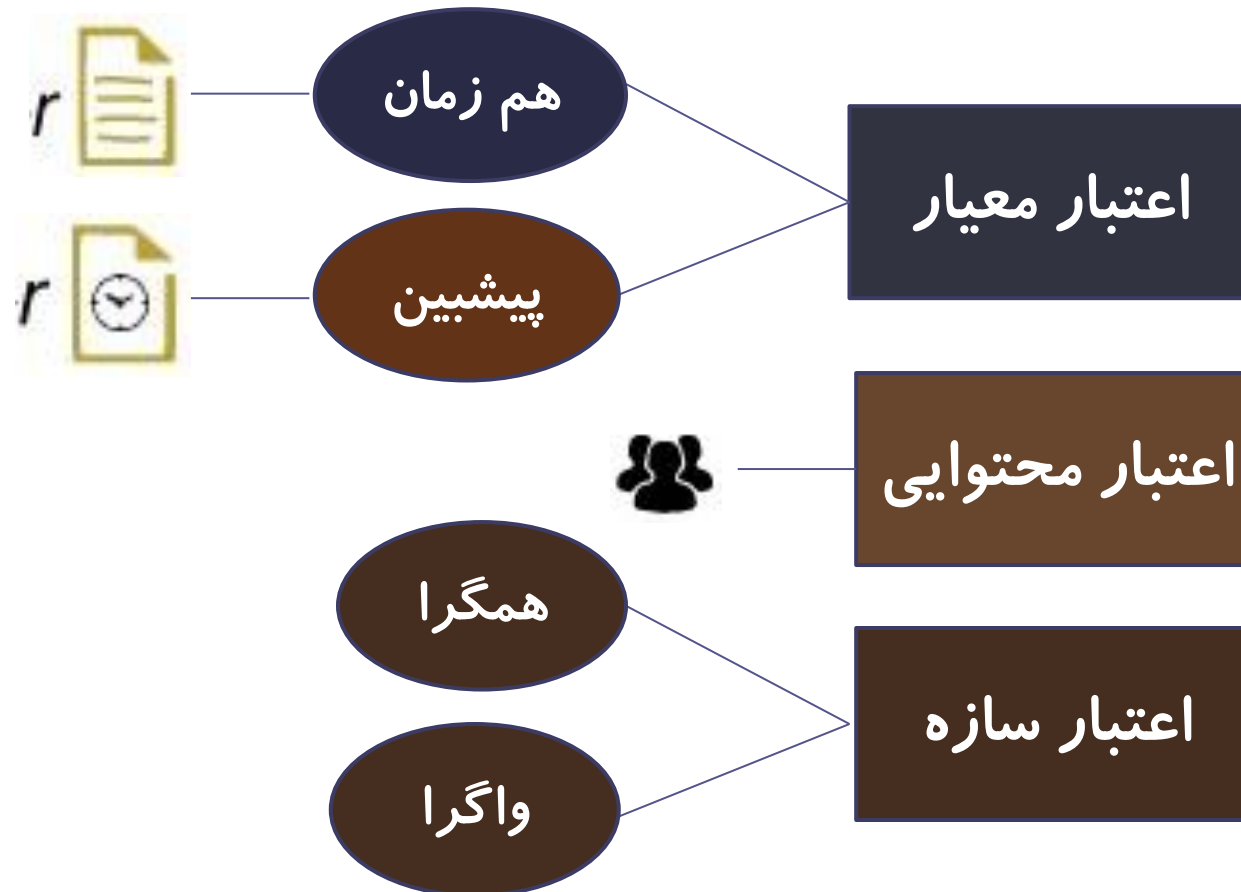
رایج ترین روش ها برای تعیین اعتبار:

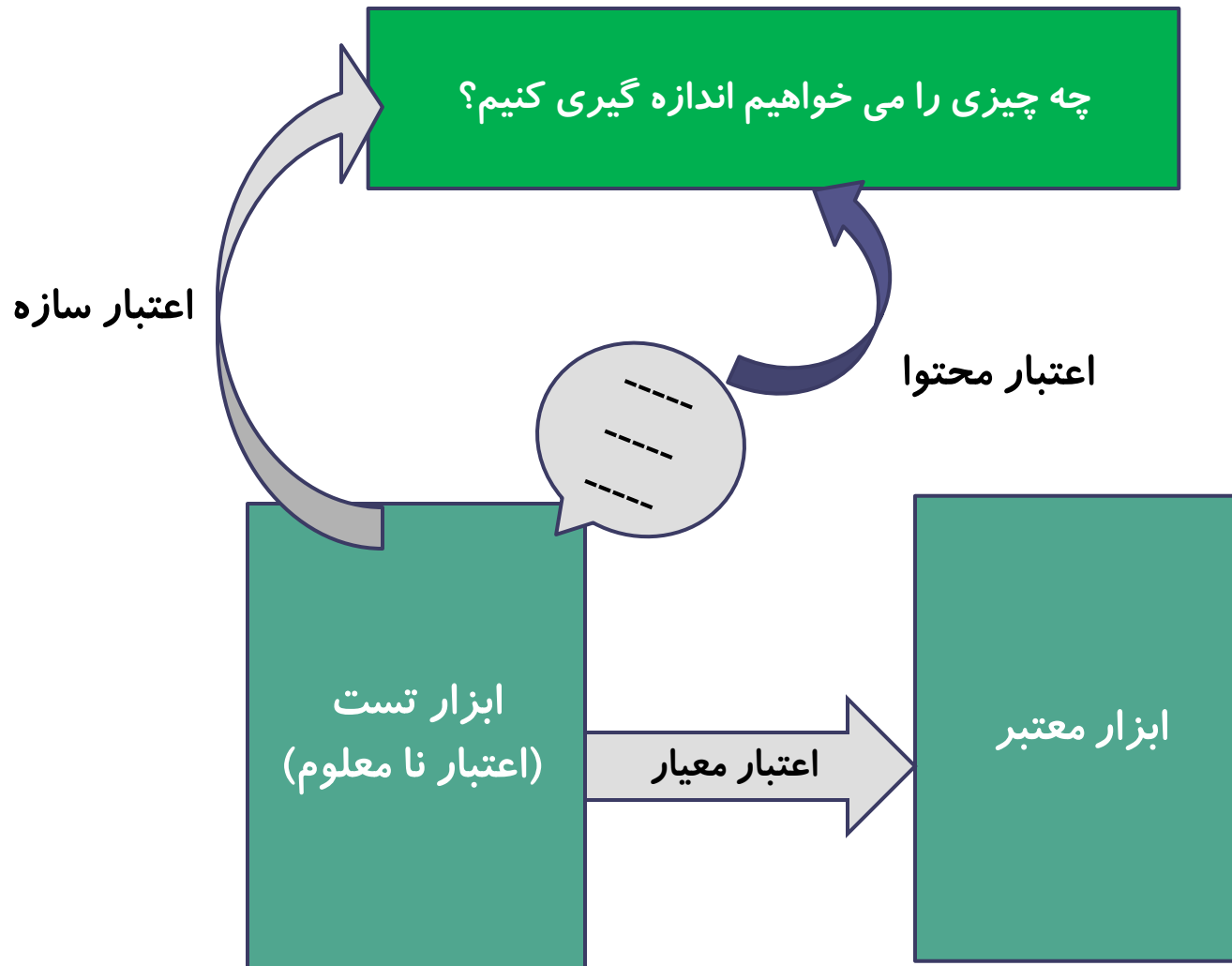


■ اعتبار محتوایی

■ اعتبار معیار

■ اعتبار سازه





اعتبار محتوا

Content validity

ارتباط ابزار یا استراتژی اندازه گیری با سازه ای که اندازه گیری می شود

- رویکرد برای تعیین اعتبار محتوا:
 - تعریف سازه و توسعه آیتم ها

اگر آیتم های تست انعکاس صحیحی از سازه نباشد،
تست فاقد اعتبار محتوی است.



فراموش نکنید

اعتبار محتوا

ارتباط ابزار یا استراتژی اندازه گیری با سازه مورد اندازه گیری است

Question	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	CVR
1	✓		✓	✓	✓	0.6
2		✓	✓	✓		0.2
3	✓				✓	-0.2
4	✓	✓		✓		0.2
5			✓		✓	-0.2
6	✓	✓		✓	✓	0.6
7	✓	✓	✓	✓	✓	1

مراحل سنجش اعتبار محتوی

مرحله ۱: جمع آوری داده از متخصصین

مرحله ۲: محاسبه نسبت اعتبار محتوی

مرحله ۲: محاسبه ایندکس اعتبار محتوی

Content Validity Ratio (CVR) = $(n_e - N/2) / (N/2)$

Content validity index (CVI). Average CVR score of all test questions

مقادیر نزدیک تر به یک بیانگر اعتبار محتوی بالا است.

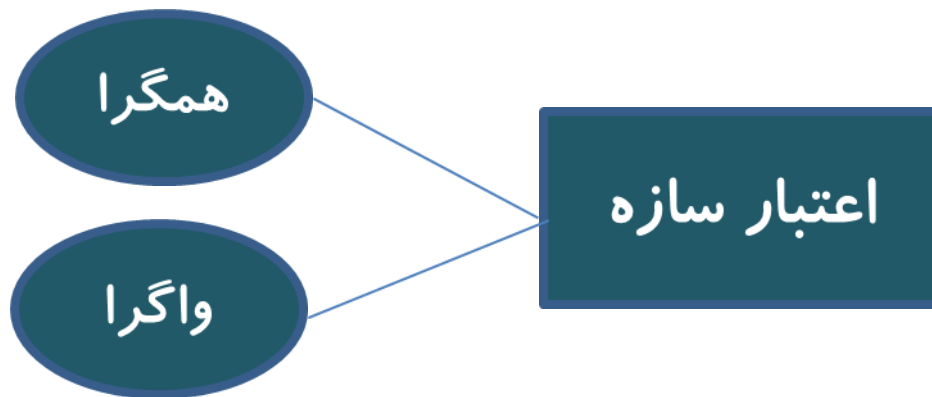
$$\text{CVI} = (0.6 + 0.2 - 0.2 + 0.2 - 0.2 + 0.6 + 1) / 7 = 0.31$$

Question	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	CVR
1	✓		✓	✓	✓	0.6
2		✓	✓	✓		0.2
3	✓				✓	-0.2
4	✓	✓		✓		0.2
5			✓		✓	-0.2
6	✓	✓		✓	✓	0.6
7	✓	✓	✓	✓	✓	1

Construct validity

اعتبار سازه

تا چه حد تست یا استراتژی اندازه گیری یک سازه یا ویژگی تئوریک را اندازه گیری می کند.



Convergent validity

اعتبار همگرا

- رابطه بین معیار مورد نظر و معیار دیگری که سازه مشابه را اندازه گیری می کند.
- این رابطه به صورت همبستگی بین دو رویکرد، یا ضریب همبستگی بیان می شود.
- همبستگی مثبت قوی بین دو معیار بیانگر اعتبار سازه است.

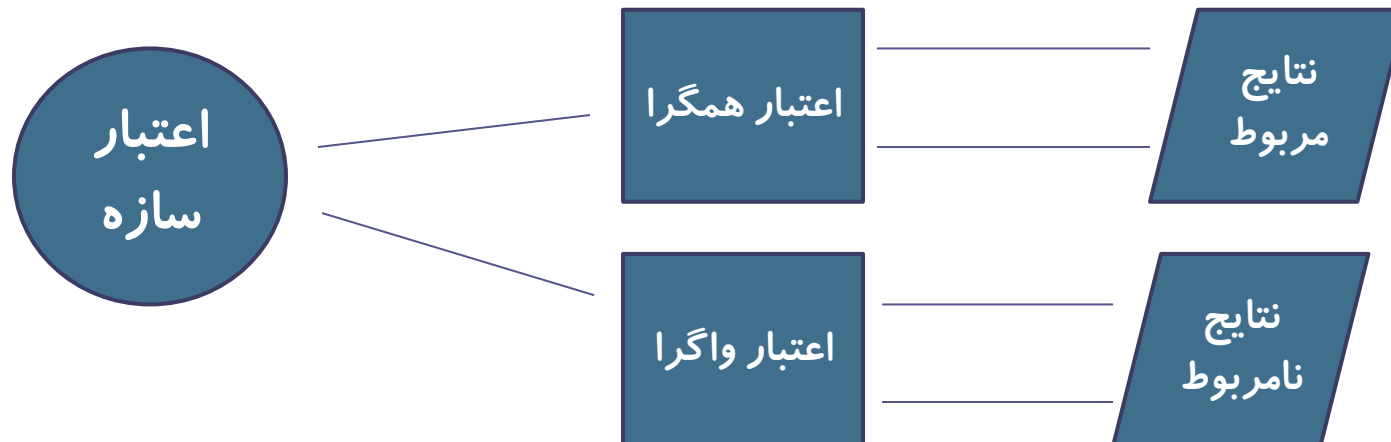
Divergent validity

اعتبار واگرا

■ نامربوط بودن دو سازه

■ نامربوط بودن معیار افسردگی با معیار شادی

■ اعتبار سازه با همبستگی منفی قوی



فراموش نکنید:

اعتبار سازه

- حدی که تست یا استراتژی اندازه گیری سازه یا ویژگی تئوریک را می سنجد.
- وجود رویکردهای مختلف برای تعیین اعتبار سازه
- اعتبار همگرا و واگرا
- حدی که اندازه گیری یک سازه معین با اندازه گیری سازه های مشابه یا مختلف همگرایی یا واگرایی دارد.



The functional rating index: reliability and validity of the Persian language version in patients with neck pain

Noureddin Nakhostin Ansari , Ronald J Feise, Soofia Naghdi, Aref Mohseni, Mahdi Rezazadeh

Affiliations + expand

PMID: 22310090 DOI: 10.1097/BRS.0b013e31824b5bde

Abstract

Study design: Psychometric testing of the Persian Functional Rating Index (PFRI).

Objective: To determine the reliability and validity of the PFRI in Persian-speaking subjects with neck pain (NP).

Summary of background data: The Functional Rating Index is a self-report questionnaire that can be used for patients with back pain or NP. The PFRI has been recently validated in patients with low back pain, whereas it is not validated in patients with NP.

Methods: One hundred patients with NP, mean age of 42 years, participated in the study; 50 patients agreed to be tested on 2 occasions during a 7-day interval for the reliability phase of the study. A visual analogue scale, the Neck Disability Index, and the Neck Pain and Disability Scale were also completed to assess validity. Fifty healthy subjects completed the PFRI for discriminative validity.

Results: Floor and ceiling effects were not observed. Independent t test showed a statistically significant difference in PFRI total scores between patients and healthy subjects supporting the discriminative validity of the PFRI ($P < 0.001$). PFRI and visual analogue scale demonstrated concurrent criterion validity, with Pearson correlation coefficients of 0.75 for test and 0.70 for retest. Construct validity was supported by a significant Pearson correlation between the PFRI and the Neck Disability Index ($r = 0.72$, $P < 0.0001$) and between the PFRI and the Neck Pain and Disability Scale ($r = 0.63$, $P < 0.0001$). Factor analysis revealed a 2-factor solution, which jointly accounted for 64.75% of the total variance. Additional factor analysis suggested 8-item PFRI as a unidimensional functional instrument for patients with NP. Internal consistency for the PFRI was high. (Cronbach α coefficients were 0.88 for test, 0.89 for retest, and 0.89 for 8-item PFRI.) Reproducibility assessed by test-retest reliability was excellent (intraclass correlation coefficient [ICC](agreement) of 0.96, $P < 0.0001$).

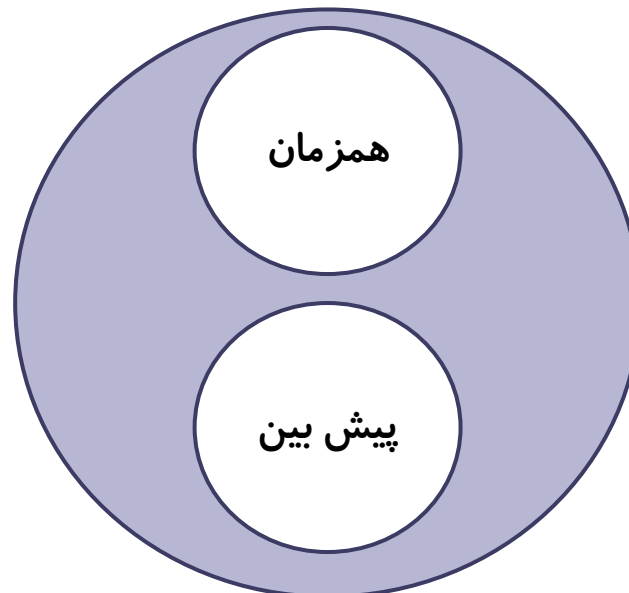
Conclusion: The PFRI is valid and reliable for use in a Persian-speaking population with NP.

Criterion validity

اعتبار معیار

✓ همبستگی با معیار بیرونی که همزمان اندازه گیری شده است به نام اعتبار همزمان

✓ همبستگی با معیار بیرونی که بعداً اندازه گیری خواهد شد به نام اعتبار پیش بین



اعتبار معیار

مثال اعتبار معیار

ابزار یا استراتژی اندازه گیری برای افسردگی

- *اعتبار همزمان*، تایید همزمان افسردگی بر اساس ضوابط تشخیصی افسردگی
- *اعتبار پیش بین*، تایید افسردگی در آینده بر اساس ضوابط تشخیص افسردگی





فراموش نکنید:

اعتبار معیار

- سنجش رابطه بین معیار مورد نظر و معیار ملاک بیرونی
- *اعتبار معیار همزمان* رابطه بین معیارهایی که همزمان اندازه گیری شده اند.
- *اعتبار معیار پیش بین* رابطه بین معیارهایی که در زمان های مختلف اندازه گیری شده اند.

› Electromyogr Clin Neurophysiol. 2008 Jan-Feb;48(1):35-41.

Neurophysiological examination of the Modified Modified Ashworth Scale (MMAS) in patients with wrist flexor spasticity after stroke

S Naghdi ¹, N N Ansari, K Mansouri, A Asgari, G R Olyaei, A Kazemnejad

Affiliations + expand

PMID: 18338533

Abstract

The Modified Modified Ashworth Scale (MMAS) is a clinical test for the measurement of spasticity. The aim of the present study was to examine the validity of the MMAS in patients with wrist flexor spasticity after stroke. 27 adult patients (14 women and 13 men) with first ever stroke resulting in hemiplegia with a mean (SD, range) age of 57.9 (11.6, 37-75) were included in the study. The outcome measures were the MMAS for the clinical assessment of spasticity, the Hslope/Mslope (Hslp/Mslp), and the H(max)/M(max) ratio for the neurophysiological evaluation. The mean of the Hslp/Mslp and the H(max)/M(max) were higher in patients with worse MMAS grades but the differences were not statistically significant. There was a significant positive correlation between the MMAS scores and Hslp/Mslp ratio as the new index of alpha motoneurone excitability or traditional index of H(max)/M(max) ratio ($r = 0.39$, $p = 0.04$). It is concluded that the MMAS to be a valid measure of spasticity after stroke.



استراتژی های اندازه گیری برای جمع آوری داده

- بعضی روش های مورد استفاده برای جمع آوری داده
- تعریف سازه های مورد نظر بر حسب متغیرهای مستقل و وابسته



چگونه متغیرهای مستقل و وابسته را اندازه گیری می کنیم؟



انتخاب استراتژی اندازه گیری به عوامل مختلفی بستگی دارد
(نظر پژوهشگر، سوال پژوهش، و طرح پژوهش)

■ پیشرفت از کل به خاص

■ مورد کلی همیشه ماهیت سوال پژوهش و متغیرهای مستقل و وابسته است.

تصمیم پژوهشگر برای اندازه گیری متغیرهای مستقل و وابسته منطبق با هدف پاسخ دادن به سوال پژوهش

ضرورت ملاحظه عوامل اختصاصی تر

اهمیت مقیاس های اندازه گیری، در چه سطحی باید سعی کنیم متغیرهایمان را اندازه گیری کنیم؟ سطح فاصله ای و نسبی اندازه گیری؟

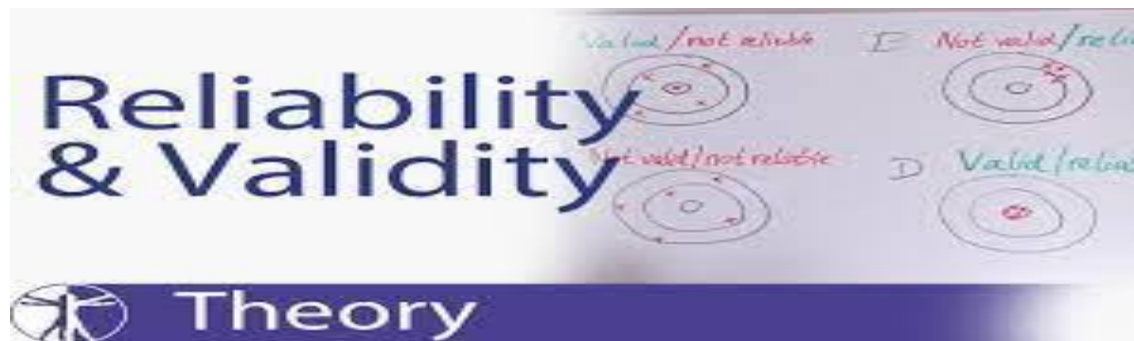
همیشه ممکن یا دلخواه نیست!



عامل دیگر: ویژگی های سایکومتریک استراتژی اندازه گیری

اهمیت پایایی و اعتبار در همه انواع اندازه گیری

صرفنظر از این که چه چیزی و چگونه اندازه گیری می شود، آن رویکرد اندازه گیری باید پایا و دارای اعتبار سازه باشد.



عامل دیگر: مناسب بودن ابزار برای جمعیت مورد بررسی



اطمینان از جمعیتی که تست برای آن ساخته شده است (مثال،
ارزیابی سطح افسردگی در سالمندان)

اگر برای جمعیت مورد نظر نباشد، استفاده از آن ابزار برای بررسی
افسردگی در جمعیت سالمند دچار افسردگی مناسب نخواهد بود.

AVAILABILITY

در دسترسی:

عامل مهم هنگام انتخاب استراتژی اندازه گیری

رویکرد دیگر:

مرور پژوهش های مربوط، دیگران چگونه سازه مورد نظر یا سازه های مشابه را اندازه گیری کردند.

عامل دیگر: هزینه

استفاده از بعضی ابزارهایی که بطور تجاری در دسترس هستند، می تواند هزینه بر باشد.

Commercially available

ممکن است لازم باشد پژوهشگر معیار یا استراتژی اندازه گیری را برای سنجش سازه های مورد نظر توسعه دهد.



دو فاکتور زمان انجام اندازه گیری و آسانی استفاده از ابزار

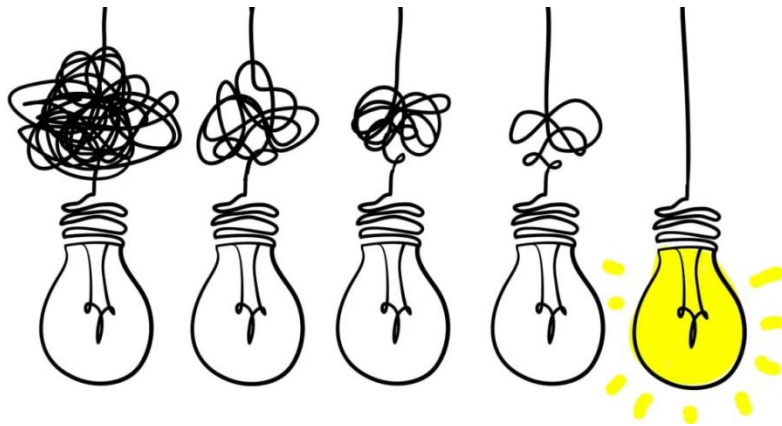
✓ مفهوم پارسیمونی

✓ انتخاب ساده ترین توضیح برای پدیده

✓ مفهوم کلیدی، سادگی است



✓ بسته به سازه، با ارزیابی طولانی تر و پیچیده تر لزوماً اندازه گیری صحیح تر نخواهد بود.



✓ افزایش احتمال

✓ اشتباهات

✓ خستگی

✓ بی توجهی

در میان پژوهشگران و افراد

از روش های ارزیابی طولانی و پیچیده غیر ضروری در هر زمان ممکن اجتناب کنید.

روش های جمع آوری داده

«بدون داده
شما فقط فرد دیگری هستید
با یک عقیده.»

ادواردز دمینگ

وجود رویکردهای مختلف زیاد برای جمع آوری داده

- ✓ سوال پژوهش و ماهیت متغیرها معمولاً تعیین کننده استراتژی اندازه گیری برای جمع آوری داده هستند.
- ✓ با ابزارهای موجود می توان مجموعه متنوعی از عوامل مربوط به تجربه انسانی را اندازه گیری کرد.
- ✓ ابزارهای موجود مورد توجه هستند چون پایایی و اعتبار آنها تثبیت شده است و نیاز به تولید و اعتبارسنجی یک ابزار جدید نیست.
- ✓ بسیاری از این ابزارها نیز داده های فاصله ای و نسبتی تولید می کنند که پیش شرط انواع معینی از آنالیزهای آماری هستند.

قبل از تلاش برای تولید یک تست جدید،
همیشه ابزارهای موجود را به عنوان روش های جمع آوری داده در
نظر داشته باشید.

یک استراتژی اندازه گیری ضعیف می تواند نتایج حتی بهترین
دیزاین پژوهشی را مخدوش کند.

راهنما: پایایی و اعتبار



استفاده از رویکردهای مختلف و معتبر جمع آوری داده



✓ تستینگ های رسمی

✓ مصاحبه

✓ ریتینگ گلوبال

✓ مشاهده

✓ معیارهای بیولوژیک





مصاحبه

رویکردی ساده و شکلی از خوداظهاری برای جمع آوری داده

✓ اطلاعات زیادی فراهم می کند.

✓ نسبتاً ارزان ، جمع آوری مجموعه وسیعی از داده بدون نیاز به آزمایش

- جمع آوری داده ها در باره تاریخچه زندگی و بیوگرافی آزمودنی ها
- برای تولید پرسشنامه خوداظهاری
- استخراج آیتم های مربوط به سازه مورد نظر

موثر بودن مصاحبه به چگونگی ساختار آن بستگی دارد.

مصاحبه باید استاندارد شده باشد:

- سوالات یکسان از همه آزمودنی ها
- پرسش به یک ترتیب

آموزش پژوهشگران انجام دهنده مصاحبه در باره نحوه اجرای صحیح آن



ضعف مصاحبه:

- نامناسب برای آنالیز آماری چون سازه را توصیف می کند
نمونه کلاسیک، مصاحبه برای استخدام افراد

✓ مصاحبه ها جزء اساسی اکثر انواع پژوهش های کیفی است.



Global rating

ریتینگ گلوبال

شکل دیگری از خوداظهاری

یک تکنیک جمع آوری داده در پژوهش ها

بر خلاف مصاحبه،

این رویکرد سازه یا متغیر مورد نظر را اندازه گیری می سنجد،
با پرسش از فرد برای نمره دادن بر روی یک طیف عددی برای
پاسخش به سوال موضوع

Global Rating

لطفا به میزان علاقه خود به وینار های پژوهش آموزی بر اساس مقیاس ۱ تا ۵ نمره دهید.

۱	۲	۳	۴	۵
بسیار کم		خنثی		بسیار زیاد

استفاده از گلوبال ریتینگ ها همچنین هنگام پرسش از افراد برای ریت

✓ حالات هیجانی

✓ علائم

✓ سطح استرس

قدرت گلوبال ریتینگ:

✓ استفاده برای مجموعه وسیعی از موضوعات و سوالات

✓ داده های فاصله ای و نسبتی

گلوبال ریتینگ فقط معیاری گلوبال از سازه است
و پیچیدگی یا جزئیات را ارزیابی نمی کند

Observation

مشاهده

روش دیگری برای جمع آوری داده

✓ بر اساس مشاهده مستقیم سازه مورد نظر، که اغلب نوعی رفتار است.

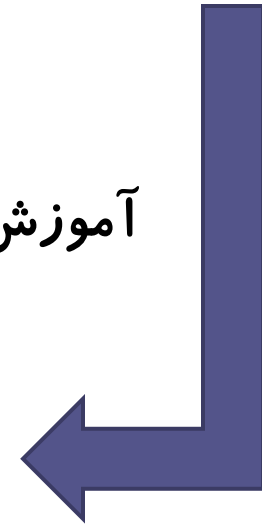
✓ این نوع رویکرد در پژوهش های مختلف بطور گسترده ای استفاده می شود.

پروسه سنجش مشاهدات باید استاندارد شود.

تعریف عملی صحیح رفتار مورد سوال

آموزش کافی افراد مسؤل جمع آوری داده برای اطمینان از صحت مشاهدات

کمک به اطمینان از اعتبار کافی و پایایی



✓ استفاده از ویدئوتیپینگ و ریتورهای متعدد برای تایید صحت مشاهدات

✓ داده ها اغلب فاصله ای یا نسبتی

معیارهای بیولوژیکی

رایج در پژوهش های پزشکی و سایکوبیولوژیکی

✓ اندازه گیری پاسخ های فیزیولوژیکی افراد

✓ ضربان قلب، تنفس، فشار خون،



www.merakldh.in





ضرورت تعریف عملی و استاندارد سازی

بررسی سطح اضطراب در پاسخ به محرک های معین



جمع آوری داده های بیولوژیکی

✓ تعریف عملی اضطراب بر اساس تعداد ضربان قلب و تنفس

✓ داده ها از معیارهای بیولوژیکی اغلب فاصله ای و نسبتی

فراموش نکنید:

استراتژی های اندازه گیری متعدد ✓

- رویکردی مطلوب
- همگرایی اندازه گیری های متعدد
- افزایش اعتماد کلی به یافته های پژوهش



