

# Metodi Statistici per la Neuropsicologia Forense

## *4a. Validità e affidabilità (teoria)*

---

Giorgio Arcara,  
Università di Padova  
IRCCS San Camillo, Venezia



# **Validità e affidabilità**

---

Validità e affidabilità sono due qualità fondamentali dei test (non solo didatticamente).



Nelle prossime slides farà una panoramica generale senza scendere nei dettagli delle formule che vedremo in un passo successivo.

Usare un test senza conoscerne validità e affidabilità può portare a grossi errori:

Per dare un'analogia (estrema) usare un test con bassa affidabilità e nessuna validità per un'interpretazione per un'interpretazione e potrebbe portare a misurare l'altezza di una persona con una bilancia che ha un margine di errore di 10 kg.

**Dal momento che i test sono strumenti di misurazione non possiamo prescindere dal conoscerne le loro qualità.**

## Validità e affidabilità (brevi definizioni)

La **validità** è la qualità di un test di misurare effettivamente il costrutto che vuole misurare.

L' **affidabilità** indica la precisione di un test.

# Come vorremmo fossero Validità e Affidabilità (ma non sono)

**Validità:** il mio test misura il mio costrutto in maniera nota e quantificabile. Es.:

- l'80% del punteggio osservato è riconducibile al costrutto di interesse
- il punteggio riconducibile al costrutto di interesse mentre  $\pm 5\%$  ad altri costrutti.

# Come vorremmo fossero Validità e Affidabilità (ma non sono)

**Validità:** il mio test misura il mio costrutto in maniera nota e quantificabile. Es.:

- l'80% del punteggio osservato è riconducibile al costrutto di interesse
- il punteggio riconducibile al costrutto di interesse mentre  $\pm 5\%$  ad altri costrutti.

**Affidabilità:** il mio test misura il mio costrutto con precisione quantificabile. Es.:

- il mio punteggio finale può essere sbagliato di  $\pm 3$
- il punteggio finale può essere sbagliato di  $\pm 5\%$

## Come vorremmo fossero Validità e Affidabilità (ma non sono)

Anche se ci sono dei criteri talvolta di soglia minima (di affidabilità o validità) non c'è una unanimità. éer considerare un test adeguato, spesso questi valori ci aiutano a distinguere test tra di loro, per scegliere il migliore (o il meno peggiore)

*Peggior Affidabilità  
e/o Validità*

*Migliore Affidabilità  
e/o Validità*

**Peggio** ————— **Meglio**



## Parlare correttamente di validità

La validità è la qualità di un test di misurare ciò che effettivamente vuole misurare. Il termine validità è sempre riferito ad un'**utilizzo** che si fa dei punteggi.

## Parlare correttamente di validità

La validità è la qualità di un test di misurare ciò che effettivamente vuole misurare. Il termine validità è sempre riferito ad un'**utilizzo** che si fa dei punteggi.

Non ha senso dire che **un test è valido**

# Parlare correttamente di validità

La validità è la qualità di un test di misurare ciò che effettivamente vuole misurare. Il termine validità è sempre riferito ad un'**utilizzo** che si fa dei punteggi.

Non ha senso dire che **un test è valido**

~~"usa questo test! è stato validato!"~~

## Un concetto da ribadire



Un test può avere validità per una certa utilizzo o per più utilizzi.

La validità non è una qualità tutto-o-nulla, ma una qualità lungo un continuum

Ad esempio ci sono evidenze che il *"Free and Cued Selective Reminding Test-it"* (Clerici et al., 2017) è un test con buona validità per misurare le capacità di memoria.

talvolta, alcuni test non riportano nessun evidenza di validità. Si tratta di strumenti in cui non abbiamo la certezza di cosa stiano misurando.

## Tipi di validità

- Validità di facciata
- Validità di contenuto
- Validità convergente/divergente (validità di costrutto)
- Validità di criterio
- Validità concorrente
- Validità ecologica

Esistono diverse classificazioni di validità e affidabilità (vedi Urbina, 2004, *Essentials of Psychological Testing*, per un approfondimento). Quelle che segue è solo una delle possibili classificazioni.

## Validità di facciata 1/2

Validità di facciata: si riferisce alla qualità di un test di essere chiaramente riconducibile al costrutto che vuole misurare. Es. un test di memoria che richiede di ricordare gli items.

Talvolta è l'unico tipo di validità che abbiamo a disposizione (ma di cui non dovremmo accontentarci.)

Data la natura prettamente qualitativa non esistono analisi per verificarla.



La plausibilità di una misurazione non è sufficiente per giudicare se effettivamente stiamo misurando quello che ci interessa.

Advances in Methods and Practices in Psychological Science  
Volume 3, Issue 4, December 2020, Pages 456-465  
© The Author(s) 2020, [Article Reuse Guidelines](#)  
<https://doi.org/10.1177/2515245920952393>



*General Article*

### **Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them**

**Jessica Kay Flake**  <sup>1</sup> and **Eiko I. Fried**  <sup>2</sup>

Validità di contenuto : la proprietà degli item di essere sufficienti ed adeguati per valutare il costrutto di interesse. Può essere valutata qualitativamente o quantitativamente (Lawshe, 1975).

Spesso non è riportata nei test o c'è solo una valutazione qualitativa. Due test che riportano validità di contenuto quantitativamente sono Abaco (Sacco et al., 2008) e APACS (Arcara & Bambini, 2016), ma non usano statistiche inferenziali.

*Esempio di test con scarsa validità di contenuto:*

Un test che vuole misurare l'abilità di lettura nella vita quotidiana e utilizza solamente parole in isolamento.

*Esempio di test con buona validità di contenuto:*

Un test che vuole misurare le abilità di guida. E simula le abilità di guida tramite un sorta di videogioco in numerose condizioni.

## Validità di costrutto (convergente/divergente) 1/4

Validità convergente-divergente (o validità di costrutto): indaga quanto il test misura effettivamente il costrutto che intende misurare valutando coerenza interna degli item, oppure correlazione con altri test. Si parla anche di validità “divergente” perché valuta anche la qualità di non correlare con test con cui non dovrebbe correlare (ad esempio un test specifico per memoria visuospatiale, dovrebbe non correlare troppo con memoria verbale).

A livello di analisi di dati. Per la coerenza degli items di un test tra loro si usa l'analisi fattoriale o l'alpha di Cronbach (per quest'ultima analisi vedi sezione affidabilità).

Per la relazione dei punteggi totali con quelli di altri test si usano spesso correlazioni. Non esistono delle regole su che valori (nell'analisi fattoriale, o nelle correlazioni) siano accettabili. Dipende dal costrutto misurato e dalle relazioni attese.

## Validità di costrutto (convergente/divergente) 3/4

Un test con buona validità di costrutto correla con test che misurano lo stesso costrutto o correla con test che misurano altri costrutti in maniera coerente con le aspettative.

*Ad esempio*, un test di memoria di lavoro con buona validità di costrutto, correla con altri test di memoria di lavoro.

Un test di comprensione di linguaggio figurato (costrutto più ampio), dovrebbe correlare con altri test che misurano lo stesso costrutto, ma in maniera moderata, potrebbe correlare anche con attenzione o altre funzioni cognitive legate.

## Validità di costrutto (convergente/divergente) 4/4

La validità di costrutto è da un lato forse la validità più importante, ma dall'altro quella più difficile da valutare se adeguata quando si indaga in relazione ad altri test che misurano lo stesso costrutto. Questo perché non esistono indicazioni di “correlazione minima” che dovrebbe avere un test con un altro per avere supporto alla sua validità di costrutto.

*Ad esempio:* se sviluppo un nuovo test di memoria, non è detto che ci sia un valore minimo di correlazione con un altro test di memoria che sia considerato come sufficiente.

Per l'analisi fattoriale o il Cronbach's alpha esistono invece principi più chiari: l'analisi deve mostrare che gli item si comportano in maniera statisticamente adeguata relativamente al costrutto (questo sarà più chiaro nella sezione di approfondimento psicometrico di queste analisi).

## Validità di criterio 1/2

Validità di criterio: La proprietà di un test di fornire risultati legati ad un criterio esterno. Quest criterio è spesso l'appartenenza ad una patologia, oppure un aspetto prognostico (lo sviluppare una patologia in futuro, il migliorare dopo un trattamento, etc.) o il punteggio ad un altro test.

La validità di criterio, se riferita ad una classificazione con un altro punteggio, è spesso espressa in termini di correlazione. Se riferita ad una classificazione in categorie è spesso espressa da valori di sensibilità/specificità (vedremo questo in maniera approfondita).

Può essere distinta in **concorrente** se il criterio è misurato nello stesso momento, o **predittiva** se lo scopo è predire.



In generale sono fornite percentuali che esprimono l'accuratezza del test nel predire il criterio. Esistono diversi standard riportati per considerare accettabile la validità di criterio, ma dovrebbe essere alta (sens/spec superiore all'80% o correlazione  $> 0.80$ )

## Validità ecologica 1/2

Validità ecologica: si riferisce alla qualità di un test di riflettere effettivamente delle abilità che hanno ripercussioni nella vita quotidiana.

(Ad esempio, che la performance deficitaria ad un test di memoria rifletta difficoltà di memoria nella vita quotidiana).

Può essere valutata tramite diverse analisi, es. Correlazioni con altri questionari o scale riferite alla vita quotidiana.

Non esistono valori soglia condivisi che un test deve avere. Nota che la validità ecologica è raramente disponibile vista la difficoltà nell'essere verificata.

Nota che **la validità ecologica è spesso trascurata** ma particolarmente importante perché spesso nelle valutazioni sono fatte inferenze implicite sull'impatto del deficit sulla vita quotidiana. Identificare un deficit di memoria è infatti relativamente importante *in sé*: quello che è rilevante è spesso capire che impatto ha questo deficit nella vita del paziente

Questo è particolarmente rilevante nel caso di valutazioni del danno subito dal paziente. Avere un deficit cognitivo che sia particolarmente invalidante è ovviamente diverso da avere un deficit che però non ha impatto sulla vita quotidiana.

Un test con dati su validità ecologica ci aiuterebbe a fare queste interpretazioni perché ci sarebbe supporto scientifico sulla relazione tra performance al test e comportamento di vita quotidiana.

## Come conoscere validità di uno strumento?

- Per conoscere la validità di un test occorre documentarsi (manuale, articolo scientifico).
- Per comprendere gli effetti della validità occorre conoscere degli aspetti (veramente base) di statistica.
- In Italia e nei test neuropsicologici c'è spesso poca attenzione alla validità. Pochi test riportano dati su validità. Il termine *validato* viene in certi casi usato per indicare che sono stati raccolti i dati normativi, un aspetto completamente diverso.



## Validità nei test italiani.

CV	CrV	DV
<input type="checkbox"/>		<input type="checkbox"/>
<input checked="" type="checkbox"/>		<input type="checkbox"/>
<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>		
	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>		

Validità nei test di screening italiani (somministrazione dal vivo). Aiello et al., 2022. Una cella senza quadratino indica che quel dato non è disponibile.

CV=concurrent validity

CrV=criterion validity

DV=divergent validity

## Un aspetto importante di validità

La validità non è qualità “immutabile” e data di un test, ma dipende da come lo si utilizza. Se il test viene utilizzato in maniera inappropriata, allora può non diventare più valido.

L'appropriatezza dell'utilizzo di un test è determinata dall'adesione alle procedure legate all'utilizzo del test stesso. Quanto più devio dalle procedure iniziali e quanto meno valido è il test.