

Metodi Statistici per la Neuropsicologia Forense

2. *Cornice teorica*

Giorgio Arcara,
Università di Padova
IRCCS San Camillo, Venezia



Cornice Teorica

"We may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or event, according to rules."

Stevens (1946)

"We may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or event, according to rules."

Stevens (1946)

Questa è una definizione estremamente popolare che tuttora domina la neuropsicologia

Scale di misura secondo Stevens

Per Stevens, quello che è importante è definire poi quale è la scala di misura, che mette in relazione oggetti con i numeri.

Scala	Descrizione	Esempio	Operazioni possibili
Nominale	Classificazione in categorie senza ordine implicito	Sesso (M/F), Colori, Nazionalità	Conteggio, moda, verifica uguaglianza/differenza
Ordinale	Classificazione con ordine, ma le distanze tra categorie non hanno significato	Posizione in una classifica, Titolo di studio	Ordinamento, mediane, percentili
Intervallo	Misura con ordine e intervalli uguali, ma senza zero assoluto significativo	Temperatura in °C, IQ	Addizione, Sottrazione, Media, Deviazione standard, z-score
Rapporto	Misura con ordine, intervalli uguali e lo zero ha un significato	Età, Scolari in anni, tempi di reazione	Tutte le operazioni aritmetiche (\times , \div , $+$, $-$), z-score, rapporti

Legenda

- *Scala*: tipo di misura secondo la classificazione di Stevens.
- *Descrizione*: caratteristiche principali della scala.
- *Esempio*: variabili reali che rientrano in quella scala.
- *Operazioni possibili*: trasformazioni matematiche e statistiche consentite dai dati.

Considerazioni su definizione di Stevens

[illegible]

Esaminiamo il MoCa:

Considerazioni su definizione di Stevens

[illegible]

Esaminiamo il MoCa:

Cosa notate nei punteggi delle parti che lo compongono?

Cosa notate nei compiti?

Considerazioni sulla definizione di Stevens 1/2

Supponiamo di avere un tavolo quadrato che abbia il lato di 40 cm, mentre un'altro tavolo abbia il lato di 42 cm. Concludereste che il primo tavolo è lungo 2 cm più dell'altro? (il cm è qui usata come unità di misura)

Considerazioni sulla definizione di Stevens 1/2

Supponiamo di avere un tavolo quadrato che abbia il lato di 40 cm, mentre un'altro tavolo abbia il lato di 42 cm. Concludereste che il primo tavolo è lungo 2 cm più dell'altro? (il cm è qui usata come unità di misura)

Supponiamo che ad un test di memoria una persona ottenga 40, mentre un'altra 42, vi sentireste di concludere che quella persona ha due unità di misura (di memoria) in più dell'altra?

Considerazioni sulla definizione di Stevens 1/2

Supponiamo di avere un tavolo quadrato che abbia il lato di 40 cm, mentre un'altro tavolo abbia il lato di 42 cm. Concludereste che il primo tavolo è lungo 2 cm più dell'altro? (il cm è qui usata come unità di misura)

Supponiamo che ad un test di memoria una persona ottenga 40, mentre un'altra 42, vi sentireste di concludere che quella persona ha due unità di misura (di memoria) in più dell'altra? o semplicemente che ha più memoria?

Considerazioni sulla definizione di Stevens 2/2

Nell'esempio appena utilizzato, inoltre, nel test di memoria, la variazione in punti sembra essere molto legata allo strumento utilizzato per valutare la memoria. Ogni test ha punteggi ottenibili diversi (es. il MoCA da 0 a 30, GEMS, un altro test per funzionamento cognitivo generale da 0 a 100).

Considerazioni sulla definizione di Stevens 2/2

Nell'esempio appena utilizzato, inoltre, nel test di memoria, la variazione in punti sembra essere molto legata allo strumento utilizzato per valutare la memoria. Ogni test ha punteggi ottenibili diversi (es. il MoCA da 0 a 30, GEMS, un altro test per funzionamento cognitivo generale da 0 a 100).

in questi esempi “+2” potrebbe avere significati molto diversi.

La misurazione in neuropsicologia clinica (e forense) praticamente sempre *assume* che ci sia una misura in **scala ad intervalli**. Ma, ad un'analisi approfondita, spesso sembrerebbe più appropriata una **scala ordinale**.

Nella scala ordinale si assume che sia solo possibile ordinare valori dal più alto al più basso ma non che si possano fare altre operazioni.

Nella scala ad intervalli si assume siano possibili operazioni come somma, sottrazione e ogni intervallo ha lo stesso significato (indipendentemente dal valore di partenza o dal range della scala). (Un'analogia è i gradi Celsius).

Misurazione e arbitrarietà della misura

L'esempio mostrato sopra illustra come nei test neuropsicologici la scala arbitraria sembra essere critica. Immaginate due test di memoria: uno ha un punteggio da 0 a 50 Un altro da 0 a 10. In entrambi i casi il valore +2 ha un significato potenzialmente diverso.

Un modo per “aggirare” questo problema (di scala arbitraria) è quello di riportare ad altre scale comparabili, come la deviazione standard oppure la percentuale di persone che ottengono quel punteggio. Queste due possibilità saranno trattate nelle lezioni su “dati normativi” e su “confrontare punteggi a test diversi”.

Un'altra possibile definizione di misurazione da filosofia e fisica

Non si misurano oggetti ma *proprietà* di oggetti:

“Sia O un certo insieme di oggetti caratterizzati dall'avere la proprietà P e sia \mathbb{R} l'insieme dei numeri Reali i quali possono essere ordinati in accordo al criterio “essere maggiore o uguale a”, si definisce misurazione:

Chiamiamo O_p l'insieme di oggetti tali che si crei un “omomorfismo” tra (O_p, \succsim) e (\mathbb{R}, \geq) .

Dove \succsim indicano *relazioni d'ordine empirico* (oggetti fisici) e \geq *relazioni d'ordine numerico*”

Boniolo e Vidali, Filosofia della scienza, p. 321

Considerazioni conclusive su misurazione 1/2

Queste considerazioni mostrano come un atto così semplice e naturale (come ottenere e utilizzare un punteggio), nasconde delle riflessioni che meritano approfondimento e che hanno influenza sul significato che assumeranno questi punteggi.

La definizione di misurazione che usiamo è un po' più permissiva rispetto a definizioni più rigide (es. quella usata fisica riportata nella slide precedente) e lascia aperte alcune riflessioni.

L'aspetto da ricordare è che nei test neuropsicologici, il più delle volte in maniera implicita, è assunta la definizione di misurazione di Stevens e che sia possibile assumere una misura in una scala ad intervalli.



Un concetto che emerge spesso quando si parla di misurazione, di test, di statistica è quello di oggettività.

Un concetto che emerge spesso quando si parla di misurazione, di test, di statistica è quello di oggettività. I test infatti vengo spesso considerati più “oggettivi”. Vale la pena quindi approfondire questo concetto.

Un concetto che emerge spesso quando si parla di misurazione, di test, di statistica è quello di oggettività. I test infatti vengo spesso considerati più “oggettivi”. Vale la pena quindi approfondire questo concetto.

Definiamo qui oggettività come una proprietà che si applica ad una valutazione, ad un test, o come *l'indipendenza dell'esaminatore dal risultato*

“L'esame neuropsicologico forense viene spesso considerato un esame oggettivo del funzionamento cognitivo. In un'ideale gerarchia delle sorgenti d'evidenza, tale esame si collocherebbe molto più in alto rispetto alle evidenze cliniche più o meno sofisticate (self-reports, informant-based reports, interviste cliniche, rating scales, ecc.), e dovrebbe pertanto essere “promosso” a rango più elevato (al pari degli esami biochimici, radiologici, ecc.). Dopo che il clinico ha esaminato il soggetto e ha formulato le sue ipotesi preliminari, chiede agli esami strumentali di confermarle, confutarle o correggerle attraverso dati oggettivi, indipendenti sia dal giudizio dell'esaminatore e dalle sue distorsioni, sia da ogni variabile estranea all'oggetto d'indagine, in particolare dalle contaminazioni dovute alla soggettività dell'esaminato.”

Copia verbatim da Stracciari, Bianchi, Sartori (2010); Neuropsicologia Forense

Il concetto di oggettività verrà ripreso più avanti, specie con *inter-rater agreement*

Definizione di test

"Un test è una procedura sistematica per l'osservazione del comportamento di una persona e per la sua descrizione con l'aiuto di una scala numerica o un sistema di categorie.

(Cronbach, 1990, p.32)

Definizione di test

"Un test è una **procedura sistematica** per l'osservazione del comportamento di una persona e per la sua descrizione con l'aiuto di una scala numerica o un sistema di categorie.

(Cronbach, 1990, p.32)

Definizione di test

"Un test è una procedura sistematica per l'osservazione del comportamento di una persona e per la sua descrizione con l'aiuto di una scala numerica o un sistema di categorie.

(Cronbach, 1990, p.32)

Definizione di test

"Un test è una procedura sistematica per l'osservazione del comportamento di una persona e per la sua descrizione con l'aiuto di una scala numerica o un sistema di categorie.

(Cronbach, 1990, p.32)

Definizione di test

"Un test è una procedura sistematica per l'osservazione del comportamento di una persona e per la sua descrizione con l'aiuto di una scala numerica o un sistema di categorie.

(Cronbach, 1990, p.32)



L'interesse non è mai nel comportamento in sé ma nel **costrutto** misurato da quel comportamento.

Costrutto: un concetto psicologico non osservabile

Cosa sono i test

I test cognitivi/neuropsicologi utilizzati in forense, sono test di performance massima (o prestazione massima)

Essi presentano alcune caratteristiche (che discuteremo nel corso). Sono diversi da questionari o test di *risposta tipica*.

Cosa sono i test

I test cognitivi/neuropsicologi utilizzati in forense, sono test di performance massima (o prestazione massima)

Essi presentano alcune caratteristiche (che discuteremo nel corso). Sono diversi da questionari o test di *risposta tipica*.

La maggior parte dei test a cui ci faremo riferimento sono quelli di performance massima perché ritenuti più rilevanti recentemente nella valutazione forense

Esempio:

Consideriamo il test Linguaggio Figurato 1.

Il paziente un punteggio finale di 10.

Cosa indica questo numero?

Esempio:

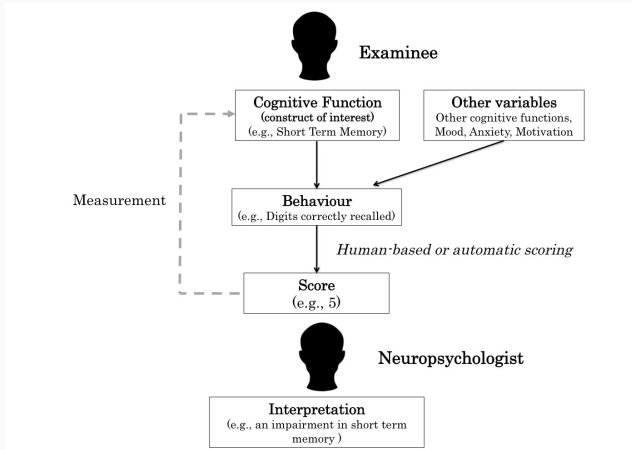
Consideriamo il test Linguaggio Figurato 1.

Il paziente un punteggio finale di 10.

Cosa indica questo numero?

Il comportamento è l'*indicatore di un costrutto* (con costrutto indico genericamente un concetto psicologico non direttamente osservabile)

Cosa sono i test



(Da Mondini, Cappelletti, Arcara, 2022)

Costrutto: Concetto psicologico non osservabile direttamente

Un esempio di test. Rivediamo l'esempio di Linguaggio Figurato 1

Classical Test Theory

La **Teoria Classica dei test (Classical Test Theory, CTT)** È un insieme di concetti di teoria di psicometria legati allo sviluppo di test psicologici. Spesso si basa su principi di misurazione come quello di Stevens

Classical Test Theory

La **Teoria Classica dei test (Classical Test Theory, CTT)** È un insieme di concetti di teoria di psicometria legati allo sviluppo di test psicologici. Spesso si basa su principi di misurazione come quello di Stevens

Fanno parte della CTT, concetti come Validità e Affidabilità.

Classical Test Theory

La **Teoria Classica dei test (Classical Test Theory, CTT)** È un insieme di concetti di teoria di psicometria legati allo sviluppo di test psicologici. Spesso si basa su principi di misurazione come quello di Stevens

Fanno parte della CTT, concetti come Validità e Affidabilità.

La maggior parte dei test neuropsicologici (clinici e forensi), implicitamente o esplicitamente utilizzano principi e assunti di CTT.

https://en.wikipedia.org/wiki/Classical_test_theory

Classical Test Theory

Il principio cardine della CTT è la seguente formula:

$$X = T + E$$

Classical Test Theory

Il principio cardine della CTT è la seguente formula:

$$X = T + E$$

X = punteggio osservato

T = punteggio vero

E = errore

Classical Test Theory

Il principio cardine della CTT è la seguente formula:

$$X = T + E$$

X = punteggio osservato

T = punteggio vero

E = errore

L'errore è indipendente dal punteggio vero.

I Limiti della Classical Test Theory

La Teoria classica dei test ha dei problemi intrinseci. Ne citiamo alcuni:

I Limiti della Classical Test Theory

La Teoria classica dei test ha dei problemi intrinseci. Ne citiamo alcuni:

- Abilità del soggetto e complessità/difficoltà del test sono confuse
- Errore e punteggio vero non sono spesso realmente indipendenti
- Problemi nella definizione di forme parallele (studieremo questo quando vedremo cambiamenti nel tempo)

I Limiti della Classical Test Theory

La Teoria classica dei test ha dei problemi intrinseci. Ne citiamo alcuni:

- Abilità del soggetto e complessità/difficoltà del test sono confuse
- Errore e punteggio vero non sono spesso realmente indipendenti
- Problemi nella definizione di forme parallele (studieremo questo quando vedremo cambiamenti nel tempo)

Consideriamo il possibile errore in un punteggio di Linguaggio Figurato 1 pari a 8, 0, oppure di 15. L'errore è veramente indipendente dal punteggio vero?

Esempio script CTT

Item Response Theory

La Item Response Theory è una famiglia di metodi più avanzati che cercano di modellare in maniera indipendente la difficoltà del task e le abilità del soggetto.

Tra i modelli di IRT i più famosi (storicamente) sono i modelli di Rasch, che permettono di modellare test in cui gli item hanno risposta sì/no.

Item Response Theory

La Item Response Theory è una famiglia di metodi più avanzati che cercano di modellare in maniera indipendente la difficoltà del task e le abilità del soggetto.

Tra i modelli di IRT i più famosi (storicamente) sono i modelli di Rasch, che permettono di modellare test in cui gli item hanno risposta sì/no.

I modelli di IRT hanno un'importante assunzione che spesso è critica per test neuropsicologici e cioè l'unidimensionalità del costrutto.

Item Response Theory

La Item Response Theory è una famiglia di metodi più avanzati che cercano di modellare in maniera indipendente la difficoltà del task e le abilità del soggetto.

Tra I modelli di IRT I più famosi (storicamente) sono I modelli di Rasch, che permettono I modellare test in cui gli item hanno risposta sì/no.

I modelli di IRT hanno un'importante assunzione che spesso è critica per test neuropsicologici e cioè l'unidimensionalità del costrutto.

Si assume che il costrutto misurato dal test sia rappresentabile lungo un'unica dimensione (non sia la combinazione di più costrutti).

Item Response Theory

Esempio formula Modelli di Rasch (la cito per completezza):

$$\Pr(X_{ni} = 1 \mid \theta_n, \beta_i) = \frac{\exp(\theta_n - \beta_i)}{1 + \exp(\theta_n - \beta_i)} \quad (1)$$

$$\Pr(X_{ni} = 0 \mid \theta_n, \beta_i) = \frac{1}{1 + \exp(\theta_n - \beta_i)} \quad (2)$$

θ = ability β = difficulty parameter

<https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory>

Item Response Theory – Rasch models

		Items:			
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
People:	1	1	0	1	1
	2	1	1	0	0
	3	1	0	0	0
	4	0	1	1	1
	5	1	1	1	0
	6	1	1	0	1
	7	1	1	1	1
	8	1	0	1	0
	9	1	1	1	1
	10	1	1	0	0

Per costruire modelli di Rasch/IRT si parte spesso da matrici di questo tipo.

Classical Test Theory vs Item Response Theory

Sebbene I modelli di IRT siano psicometricamente più rigorosi, in questo corso parlerò praticamente sempre di CTT perché la quasi totalità. Di test disponibili per Neuropsicologia Clinica e forense sono sviluppati secondo questo principio (e non con IRT).

Classical Test Theory vs Item Response Theory

Perché non ci sono test sviluppati IRT, se sono più rigorosi?

Classical Test Theory vs Item Response Theory

Perché non ci sono test sviluppati IRT, se sono più rigorosi?

Un misto di aspetti teorici e praticità

Classical Test Theory vs Item Response Theory

Perché non ci sono test sviluppati IRT, se sono più rigorosi?

Un misto di aspetti teorici e praticità

- Sviluppare test con IRT è più costoso (e può andare più facilmente male)
- spesso I costrutti neuropsicologici per forense violano quasi per definizione l'unidimensionalità.
- talvolta ad una diversa procedura, comunque IRT e CTT possono portare a risultati molto simili (Hula et al., 2006)

Classical Test Theory vs Item Response Theory

L'assunto di unidimensionalità, assume che:

- gli Item possano essere ordinati secondo la loro difficoltà e che
- la probabilità di rispondere correttamente ad un Item rifletta (più o meno) alla sua difficoltà: più un item è difficile meno probabile è rispondere correttamente.

Questo è facile pensarlo per test che includano item in cui è facile gradare la difficoltà (es. esercizi di matematica), ma con un test neuropsicologico?

Classical Test Theory vs Item Response Theory

MONTREAL COGNITIVE ASSESSMENT (MOCA)
- ITALIA -

NOVITÀ: _____
Scatola: _____ Data di nascita: _____
Sesso: _____ DATA: _____

VISUOSPAZIALE / ESECUTIVO		Copia		Disegni un orologio (fanti e dieci)		Punti	
						.../5	
DENOMINAZIONE 		.../3					
MEMORIA Leggere le serie di parole. Il soggetto deve ripetere. Fare 3 "richiami" dopo il 1° e 3°. 1° prova: _____ 2° prova: _____ 3° prova: _____		Faccia Veluto Chessa Margherita Rosso				0 punti	
ATTENZIONE Leggere la serie di cifre. Il soggetto deve ripetere. Il soggetto deve ripetere in ordine inverso. 1 2 3 4 5 5 4 3 2 1		.../2					
Leggere la serie di lettere. Il soggetto deve dare un oggetto con la mano sul tavolo ad ogni lettera "X". Si punterà 3 punti. [] F B A C M N A A G H L B A F A H D E A A G A M O F A A B		.../1					
Sottrazione di 7 partendo da 100 per 5 volte. [] 80 [] 60 [] 70 [] 72 [] 45		.../3					
LINGUAGGIO Ripetere: Si sono oggi dobbiamo aiutare Giovanni. Il gatto si nascondeva sempre sotto il divano quando c'erano cani nella stanza.		.../2					
Puntare (1° e 2° minuto): nonni il maggior numero possibile di parole che iniziano con la lettera "P". [] (1° e 2° minuto)		.../1					
ASTRAZIONE Similitudine tra per es. banana / arancio e frutto. [] Il ferro / l'orologio. [] Il cane / il pesce.		.../2					
RICORDO DEFERITO Deve ricordare la parola SENZA AUTO.		Faccia Veluto Chessa Margherita Rosso [] [] [] [] [] []		Punti vale per memoria SENZA AUTO		.../5	
Opzionale: AUTO. Categoria: Senza. Scelta multipla.							
ORIENTAMENTO [] Data [] Mese [] Anno [] Giorno [] Luogo [] Città		.../6					
© Z. Nasreddine. Traduzione a cura di A. Piani, C. Tulipani, M. Neri. Versione: 28 Luglio 2006 www.mocaitalia.org		Normale: 28 / 30 Aggiungere 1 punto se 5-12 anni di istruzione		TOTALE .../30			

È facile gradare gli item del moca per difficoltà?

Classical Test Theory vs Item Response Theory

Un aspetto teoricamente importante è che il neuropsicologo deve sapere che ogni singola prova coinvolge l'utilizzo di più di una funzione cognitiva. Per esempio, un 'Test di memoria' implica un certo grado di attenzione e non può prescindere dalle abilità di comprensione verbale; le prove di attenzione richiedono l'utilizzo di strategie cognitive che ottimizzano la prestazione o abilità di percezione visiva; il Disegno dell'orologio richiede abilità di pianificazione, capacità prassiche oltre a capacità di rappresentazione. Pertanto a ogni test non corrisponde un'unica e specifica funzione, anche se è evidente che ogni prova coinvolge maggiormente un'abilità cognitiva. In generale, possiamo dire che le prove dell'ENB-2, nel loro complesso, valutano aspetti diversi di processi cognitivi, quali: accesso e selezione lessicale, comprensione verbale, astrazione verbale, attenzione selettiva, attenzione divisa, attenzione sostenuta e attenzione alternata, funzioni esecutive di controllo, memoria a breve termine, memoria a lungo termine, memoria di lavoro, abilità prassiche, riconoscimento visivo, ragionamento logico.

I costrutti misurati tramite i test neuropsicologici

Costrutto: Concetto psicologico non osservabile direttamente

Quali sono i tipici costrutti misurati dai test neuropsicologici?



I test neuropsicologici misurano costrutti di diverso tipo

fl costrutti misurati tramite i test neuropsicologici sono a diversi livelli

Abilità funzionali

Costrutti più generici

Funzioni cognitive

1) I test neuropsicologici misurano funzioni cognitive

Esempi

Il digit span misura la (capacità della) memoria a breve termine.

Il test di memoria con interferenza (ENB-3) misura la (capacità della) memoria di lavoro.

Il TMT-A misura la (capacità di) ricerca visuo-spaziale, di attenzione selettiva e di velocità psicomotoria.

da ENB-3 (Mondini et al., 2011)

2) I test neuropsicologici misurano costrutti più generici

Esempi

Il Fab-it (Appollonio et al., 2005) misura il generale stato delle funzioni esecutive

Il SET (Dodich et al., 2015), nel punteggio globale misura l'abilità di teoria della mente.

Linguaggio Figurato 1 in APACS (Arcara & Bambini, 2016), il punteggio indica la capacità di andare oltre il significato letterale e di comprendere il linguaggio figurato.

2) I test neuropsicologici misurano costrutti più generici

Esempi

Driving Scenes Test (Brown et al., 2005)

NADL-F (Arcara et al., 2017) <https://osf.io/d9jng/>

I test ecologici molto utili per risposte concrete sulla vita quotidiana ma possono essere lunghi o poco disponibili

I costrutti misurati tramite i test neuropsicologici sono a diversi livelli

Abilità funzionali (es. Abilità di guida)

Costrutti più generici (es. Efficienza funzioni esecutive)

Funzioni cognitive (es. Memoria di lavoro)

I costrutti misurati tramite i test neuropsicologici sono a diversi livelli

Il fatto che I costrutti siano a gerarchie diverse implica che il costrutto misurato da un test potrebbe essere incluso all'interno di un costrutto misurato da un altro test

Questo aspetto verrà ripreso nel punto in cui discuteremo dell'interpretazione dei test.

Altre tipologie di test

I test neuropsicologici sono il più delle volte test di performance massima cioè test in cui al soggetto è chiesto di fare il meglio possibile.

Talvolta vengono usate anche scale funzionali o questionari (non di prestazione massima).

La **IADL** (Lawton & Brody, 1970) è una scala dell'autonomia nella vita quotidiana, non è un vero e proprio test.

Il **COAST** (Bambini et al., 2016) è una scala per la valutazione dell'efficacia comunicativa dal punto di vista del paziente o del caregiver.

Test di personalità, di umore, etc. **NON** sono test di prestazione massima

Perché è utile misurare un certo costrutto?

Perché è utile misurare un certo costrutto?

Talvolta ci si dimentica il *perché* si vuole misurare un certo costrutto e perché potrebbe essere utile, sia a livello clinico, sia a livello forense.

¹La validità ecologica verrà trattata nelle slides 4 Validità e Affidabilità

Perché è utile misurare un certo costrutto?

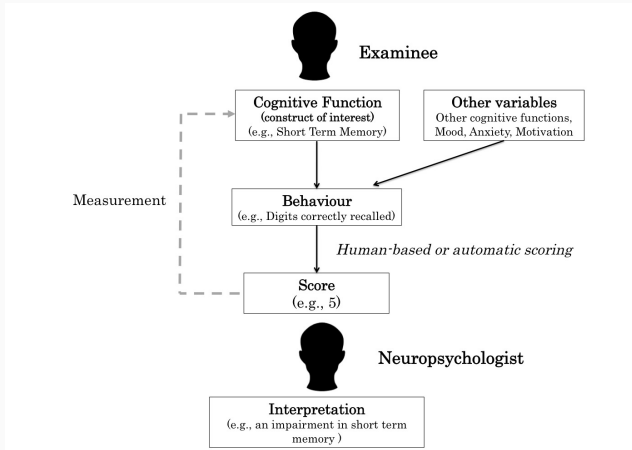
Talvolta ci si dimentica il *perché* si vuole misurare un certo costrutto e perché potrebbe essere utile, sia a livello clinico, sia a livello forense.

C'è spesso un'assunzione di fondo: che la valutazione di aspetti cognitivi (anche di base) ci può dire qualcosa su come funziona quotidianamente una persona e quindi darci informazioni sul suo comportamento in situazioni al di fuori della valutazione

Si assume che ci sia una **validità ecologica**¹ di uno specifico test (il fatto che sia andato male in questo test implica che nella vita quotidiana c'è un effetto, che ha rilevanza per il mio caso forense)

¹La validità ecologica verrà trattata nelle slides 4 Validità e Affidabilità

Assumere che i risultati del test ci dicano qualcosa sulla vita quotidiana non è sbagliato in sé, ma non dimentichiamo che (a meno che di validità ecologica comprovata da dati scientifici) questa è un'interpretazione.



(Da Mondini, Cappelletti, Arcara, 2022)

Costrutto: Concetto psicologico non osservabile direttamente