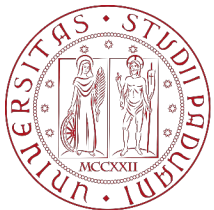


Metodi statistici per la Neuropsicologia Forense

A.A. 2024/2025

Giorgio Arcara

IRCCS San Camillo, Venezia
Università degli Studi di Padova





Simulazioni e validità di performance



Simulazioni e validità di performance

Nel contesto degli studi di forense si parla spesso di validità di performance per valutare se una specifica valutazione è da considerare *valida*

Nota che abbiamo parlato di Validità come una proprietà dei test, mentre questa è una proprietà della valutazione

In questa specifica istanza di misurazione, sto valutando ciò che intendo valutare?

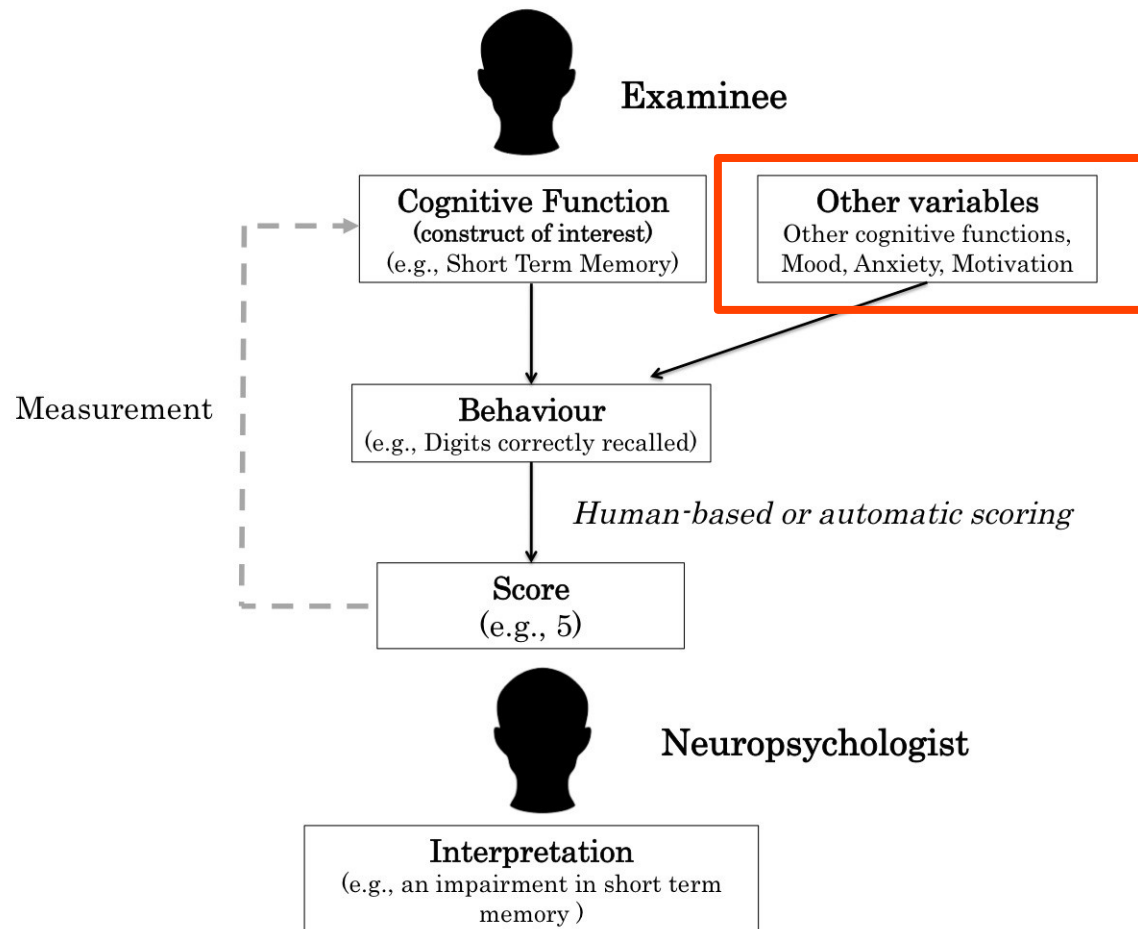


Simulazioni e validità di performance

Esistono strumenti per valutare se una performance è da considerarsi valida (**Performance Validity Test, PVT**)

Sono anche noti come **Effort Tests** (test dello sforzo)

Ricordiamo che I test cognitivi sono test di *performance massima* (Vedi slides cornice teorica). Basta impegnarsi di meno (o “sforzarsi di meno” e si invalida il risultato del test.



*Non impegnarsi adeguatamente inficia la relazione tra
costrutto da misurare e punteggio*

(Da Mondini, Cappelletti, Arcara, 2022)



Una nota su simulazione

Nota che considerare la simulazione come una condizione sì/no (o sto simulando o non sto simulando) è probabilmente una semplificazione (Nies & Sweet, 1994)

È più ragionevole pensare che si la performance possa essere lungo un continuum tra la performance genuina e quella con simulazione.

Entrano inoltre in gioco alcuni aspetti (es. Aspetti psicologici inconsci) che potrebbero portare a performance deficitarie in assenza di deficit cognitivo/danno.

Nelle slides non approfondiremo questi aspetti (molto importanti) perché più teorici e ci concentreremo sugli aspetti statistici.



Malingering, Symptom Validity Assessment Performance Validity Testing

Ci sono dei concetti importanti legati che è bene esplicitare e termini che torneranno nelle slides successive.

Malingering (Simulazione): la produzione intenzionale di problemi fisici o psicologici falsi o grossolanamente esagerati. La motivazione del malingering è solitamente esterna (ad esempio, evitare il servizio militare o il lavoro, ottenere un compenso finanziario, eludere un'azione penale o procurarsi farmaci).

Symptom Validity Assessment (SVA): l'insieme di tecniche (interviste, test, raccolta evidenze) per valutare la credibilità di sintomi (cognitivi e fisici) manifestati da un individuo.

Performance Validity Testing (PVT): l'insieme di tecniche statistiche e metodologiche per assicurarsi che il paziente stia affrontando in maniera adeguata la situazione di valutazione (sforzo adeguato, motivazione adeguata).



Come identificare simulazione

In generale si utilizza (come per cut-offs per deficit/danno) un principio probabilistico

Si costruiscono test in maniera tale che la performance mi dimostri che è probabile che la persona stia mentendo.



Le principali strategie per identificare Malingering


- 1) **Pattern di risposte inverosimili anche rispondendo in maniera casuale.** Si possono usare in test a scelta forzata e vedendo se la performance osservata è poco probabile anche rispondendo in maniera random (e quindi suggerendo menzogna intenzionale)
- 2) **Test molto facili persino per pazienti deteriorati.** Valori normativi ottenuti da popolazioni cliniche per mostrare che persino in popolazioni cliniche anche molto gravi, la performance attesa non è così bassa.
- 3) **Test sviluppati con validità di criterio:** con gruppi di persone a cui era chiesto esplicitamente di fingere un deficit, per avere un modello di “simulazione”. Queste persone possono essere sani oppure pazienti/categoria target (es. popolazioni psichiatriche)

1) Pattern di risposte inverosimili, anche rispondendo casualmente.

$$x: P(o < x | R) = k_s$$

Questa strategia può essere applicata solo nel caso di test a scelta forzata (es. Immagine di destra o immagine di sinistra?).

In questo caso la probabilità di ottenere il pattern osservato se la scelta fosse random (50% per ciascuna possibilità) k_s indica un valore di probabilità molto piccolo (in genere è il solito 5%).



1) Pattern di risposte inverosimili, anche rispondendo casualmente.

Il Test of memory Malingering (TOMM)

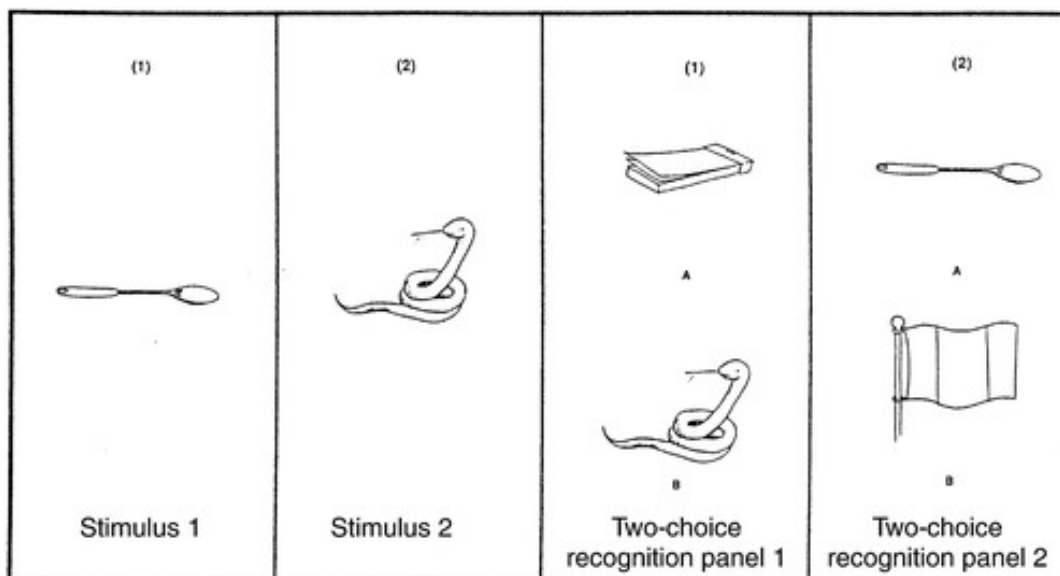
Il test of Memory Malingering è il test più comune con cui è possibile identificare questo tipo di disturbi. Si basa su un assunto molto semplice. Nel caso di una scelta multipla a due possibilità è molto facile modellare la probabilità di un certo numero di risposte casuali.

Consiste di due fasi: in una fase di fanno riconoscere degli item per ricordarli (questa viene ripetuta due volte nella versione originale del TOMM).

Nella fase test una gli item presentati nella fase di test sono presentati in coppia con un distrattore. Per ogni coppia l'esaminato deve indicare quale era l'item che aveva visto in precedenza.

1) Pattern di risposte inverosimili, anche rispondendo casualmente.

Il Test of memory Malingering (TOMM)



Esempio di stimolo del TOMM

1) Pattern di risposte inverosimili, anche rispondendo casualmente.

Il Test of memory Malingering (TOMM)

L'assunto è semplice. Un simulatore tenderà ad indicare volontariamente più spesso del caso l'immagine sbagliata. Questo però di fatto suggerisce che ha riconosciuto quale è l'immagine giusta.

In caso di scelta multiple è semplice modellare la probabilità attesa di un certo pattern di risposte (dato un certo numero di trial) usando come riferimento la distribuzione binomiale.

Es. su 50 item anche rispondendo in maniera totalmente casuale ci sono meno del 5% di probabilità di rispondere a meno di 19 item assumendo che c'è il 50% di probabilità di selezionare l'item corretto anche casualmente.

Vedi codice



1) Pattern di risposte inverosimili, anche rispondendo casualmente.

Il Test of memory Malingering (TOMM)

Vantaggi

- Il TOMM è un test molto semplice e di fatto non richiede dati normativi. Ma solo modellare la probabilità di risposte casuali.
- Si è inoltre mostrato molto efficace nell'identificare simulatori (buona sensibilità e specificità)

Svantaggi

- Se si conosce o si intuisce come funziona il TOMM è facile aggirarlo, rispondendo in maniera mista o corretta e poi andare male in altri test.
- Una parte potrebbe preparare il cliente a test di questo tipo.



2) Test molto facili persino per pazienti deteriorati.

$$x: P(o < x | D) = k_s$$

Dove k_s indica un valore di probabilità molto piccolo (es. 5%) persino in pazienti molto deteriorati (D).

2) Test molto facili persino per pazienti deteriorati.

Il Word Memory Test (WMT)

Il Word Memory Test è un test di memoria di coppie di parole (Green et al., 1999)

È diviso in più fasi.

Nella fase 1 sono presentate 20 coppie di parole (cane/gatto, pesce/pinna)
Presentate per 6 secondi nello schermo a coppia per due volte.

Immediatamente dopo sono presentate 40 coppie, di cui una sola parola era presentata originariamente e l'altra è un distrattore (es. cane/coniglio; gatto/topo).

Dopo 30 min, senza che fosse stato detto, sono presentate altre 40 coppie con altre 40 distrattori (es. cane/ratto, etc.)

Spesso in questi test (così come per ogni test di malingering) si stressa che il test è difficile, per indurre la persona a credere che una persona con deficit avrebbe una performance bassa (quando in realtà non è vero).



2) Test molto facili persino per pazienti deteriorati.

Vantaggi

- È abbastanza facile ottenere cut-off in questa maniera

Svantaggi

- Soffre dello stesso problema di tutti i cut-off di normalità: si fa un ragionamento inverso, ma non c'è sensibilità o specificità nell'identificare un simulatore.
- Per questo molti test (incluso il RDS) sono comunque validati secondo il metodo che segue (validità di criterio).

3) Test sviluppati con validità di criterio.

In questi test valgono le considerazioni per sensibilità e specificità

$$y: P(o < y | C) = k \quad \text{Sensibilità}$$

$$y: P(o \geq y | N) = j \quad \text{Specificità}$$

Dove y è il cut-off di discriminazione, o è il punteggio osservato. C è l'appartenenza al gruppo di persone con Condizione di Interesse (es. Simulatori) e N appartenenza al gruppo di non simulatori

k = Sensibilità, j = Specificità

NOTA: questa slides è idenfica alla slide su Sensibilità e specificità presentata in precedenza

3) Test sviluppati con validità di criterio.

Esistono vari modi con cui è definita la condizione di interesse *simulatori*.

- 1) si creano gruppi per condizione di interesse chiedendo a persone sane di simulare (anche dando incentivo economico in caso di successo per aumentare realismo) e sono confrontati con altri sani o con pazienti che è inverosimile che simulino.
- 2) si identifica presenza condizione di interesse (simulatori vs non simulatori) sulla base di applicazione di altri strumenti di PVT (anche combinazione di più strumenti). In tal caso il risultato agli altri strumenti di PVT diventa il gold-standard.
- 3) si usano gruppi in cui è estremamente verosimile che ci sia simulazione *probable malingering* cioè gruppi patologici per cui c'è già forte sospetto di simulazione rispetto ad altri in cui questo non c'è. Un classico esempio è confrontare pazienti che stanno facendo riabilitazione (bassa probabilità di essere simulatori) con pazienti che hanno chiesto valutazione per riscuotere assicurazione e che hanno deficit cognitivi lamentati molto gravi rispetto a funzionamento quotidiano apparentemente alto (alta probabilità di essere simulatori). In tal caso, l'appartenza a un gruppo rispetto a questi criteri è il gold standard.

3) Test sviluppati con validità di criterio.

3) si usano gruppi in cui è estremamente verosimile che ci sia simulazione (es. gruppi patologici per cui c'è già forte sospetto/certezza rispetto ad altri in cui questo non c'è (es. Gruppo di persone in cui viene fatta riabilitazione e nessuna possibilità di guadagno economico da deficit/danno)).

Esempio di approfondimento: Slick (1999) identifica questi criteri

In determining the presence of MND, the case must be evaluated on the basis of four criteria:

- (A) presence of substantial external incentive,
- (B) evidence from neuropsychological testing,
- (C) evidence from self-report, and
- (D) behaviors meeting the necessary B and C criteria are not fully accounted for by psychiatric, neurological, or developmental factors.

All diagnoses of malingering require the presence of an external incentive (Criterion A).



3) Test sviluppati con validità di criterio.

Il Reliable Digit Span

Il Reliable Digit Span non è altro che la somma del Digit Span Avanti e Digit Span Indietro.

Si basa sul fatto che le persone possono sovrastimare la difficoltà di questo test e avere una performance di fatto molto (troppo bassa), simile a 2. Spesso però è sono riportare validità di criterio rispetto alla possibilità di distinguere tra gruppi di pazienti.

3) Test sviluppati con validità di criterio.

Il Reliabile Digit Span

TABLE 5
Cutting Scores, Sensitivity, and Specificity for the Control Versus Probable Malingering Groups

<i>Reliable Digit Span Cutoff</i>	<i>Sensitivity</i>	<i>Specificity</i>
5	21	100
6	38	97
7	67	93
8	88	80

TABLE 6
Predictive Power of the Reliable Digit Span at Selected Base Rates

<i>Base Rate</i>	<i>Cutoff</i>			
	5	6	7	8
Positive Predictive Power				
.5	1.0	.95	.92	.81
.4	1.0	.88	.87	.74
.3	1.0	.85	.80	.65
.2	1.0	.80	.68	.53
.1	1.0	.57	.54	.33
Negative Predictive Power				
.5	.56	.61	.75	.87
.4	.65	.70	.81	.91
.3	.74	.78	.87	.93
.2	.83	.87	.91	.97
.1	.92	.94	.97	.99

Mathias et al. 2002 Detecting Malingered Neurocognitive Dysfunction Using the Reliable Digit Span in Traumatic Brain Injury. *Assessment*.

3) Test sviluppati con validità di criterio.

Il Reliabile Digit Span

TABLE 6
Predictive Power of the Reliable
Digit Span at Selected Base Rates

Base Rate	Cutoff			
	5	6	7	8
Positive Predictive Power				
.5	1.0	.95	.92	.81
.4	1.0	.88	.87	.74
.3	1.0	.85	.80	.65
.2	1.0	.80	.68	.53
.1	1.0	.57	.54	.33
Negative Predictive Power				
.5	.56	.61	.75	.87
.4	.65	.70	.81	.91
.3	.74	.78	.87	.93
.2	.83	.87	.91	.97
.1	.92	.94	.97	.99

Ricordiamo che Positive Predictive Power (o Value) indica:

Quale percentuale dei positivi sarà effettivamente un simulatore?

Questo dipende da quanti “simulatori” ci aspettiamo ci siano nella popolazione.

Avere questi valori può aiutarci nell’interpretare i risultati.

Es. se individuo ottiene valore < 5. Potremmo dire “anche ipotizzando un base rate basso pari a .1 (cioè 10% di simulatori nella popolazione). Una performance inferiore a 5 ha 100% sensibilità e 92% specificità)



3) Test sviluppati con validità di criterio.

Vantaggi

Test sviluppati con validità di criterio permettono di avere valori di sensibilità/specificità PPV e NPV che aiutano molto nell'interpretazione.

Svantaggi

In questi test il limite principale è l'adeguatezza del gruppo di "Simulatori".

In alcuni casi si *assume siano simulatori*, es da altri test.

In caso in cui si chiede esplicitamente di simulare, può essere una scelta poco ecologica: nella simulazione legata allo sviluppo del test la posta in gioco sarà inevitabilmente più bassa (Es. Uno studente che vince 10 € in più se finge bene non può essere paragonato ad un paziente che finge un deficit per avere un maggiore incasso da assicurazione).

Altri test comunemente usati per simulazione

Structured Inventory of Malingered Symptoms (SIMS)

Esistono anche altri test che non sono di *performance massima* che sono legati allo studio della simulazione. Un esempio è lo **Structured Inventory of Malingered Symptoms (SIMS)** che è un questionario con risposta si/no.

È diviso in varie item che sono divisi in scale per identificare simulazione di disabilità intellettive, di disturbi fisici, oppure di disturbi psichiatrici. Si ottiene anche un punteggio totale

Esempi

<i>Negli ultimi giorni faccio fatica a dormire</i>	<i>SI/NO</i>
--	--------------

<i>La mia memoria è talmente peggiorata che per alcuni giorni non riesco a ricordare nulla</i>	<i>SI/NO</i>
--	--------------

Altri test

Structured Inventory of Malingered Symptoms (SIMS)

Bello studio originale SIMS ha evidenza di validità di criterio usando una serie di gruppi di partecipanti a cui era chiesto di fingere specifici deficit (es. Psichiatrici, cognitivi, etc.)

In particolare elevata sensibilità, specificità etc. (Smith & Burger 1997)

Detection of Malingering

Table 1
Classification Rates for Individual Scales: All Groups Combined

Scale	Sensitivity %	Specificity %	Efficiency %	χ^2
Psychosis (P)	79.5	72.7	78.6	38.705*
Amnesic disorders (AM)	88.3	90.9	88.7	101.649*
Neurologic impairment (N)	85.4	75.8	84.0	59.765*
Affective disorders (AF)	74.6	72.7	74.4	29.334*
Low intelligence (LI)	85.4	51.5	80.7	24.396*
Total score	95.6	87.9	94.5	147.668*
F	94.6	81.8	92.9	123.826*
F-K	91.2	69.7	88.2	73.971*
Faking bad (FB)	82.4	69.7	80.7	41.440*
Malingering retardation (MgR)	73.7	84.8	75.2	43.085*
Malingering insanity (MgI)	79.0	72.7	78.1	37.638*

* $p < .001$.



Considerazioni conclusive (1/2)

L'identificazione di simulatori si basa sugli stessi principi statistici utilizzati per identificare deficit/danno o altre condizioni di interesse (patologia).

In questo caso il test è spesso costruito in maniera da indurre chi è malintenzionato ad avere un certo pattern di risposte che è *improbabile* e che un paziente con un reale deficit (anche grave) non commetterebbe.

Molto spesso i test di simulazione sono associati a sensibilità e specificità nell'identificare i simulatori e anche a PPV (Positive Predictive value) o NPV (Negative Predictive Value) secondo vari scenari.

Il tema della simulazione è complesso anche perché (a fini statistici) spesso si usa la distinzione netta simulazione/non simulazione, quando è verosimile che siano diversi gradi di simulazione.



Considerazioni conclusive (2/2)

Anche per I test specifici per le simulazioni le stesse considerazioni fatte per ogni altro aspetto statistico.

Sono preferibili campioni ampi per avere delle stime più solide dei cut-off (campioni più rappresentativi della popolazione)

Sono presenti I limiti del confronto con un singolo campione normativo che potrebbe non essere adeguato per i nostri scopi (es. Per età, scolarità). Sarebbe ideale avere sempre solo persone *il più simili possibili* all'individuo da valutare.

Ogni altra considerazione rimane pertinente (es. Meglio norme paese-specifiche, norme recenti, etc.)