

Bioingegneria della Riabilitazione

3.3 La misura in riabilitazione

Misurare cosa?

Misurare come?

Lorenzo Chiari, PhD

Dipartimento di Ingegneria dell'Energia Elettrica e
dell'Informazione

Alma Mater Studiorum – Università di Bologna

lorenzo.chiari@unibo.it



Proprietà di uno strumento di misura

La **misurazione** è quel procedimento che permette di ottenere la descrizione **quantitativa** (=numerica) di una grandezza fisica (la **misura**) che è per sua natura, universale, lineare, specifica.



Alcune sorgenti di rumore

Casuali
Sistematiche

Rumore
ambientale

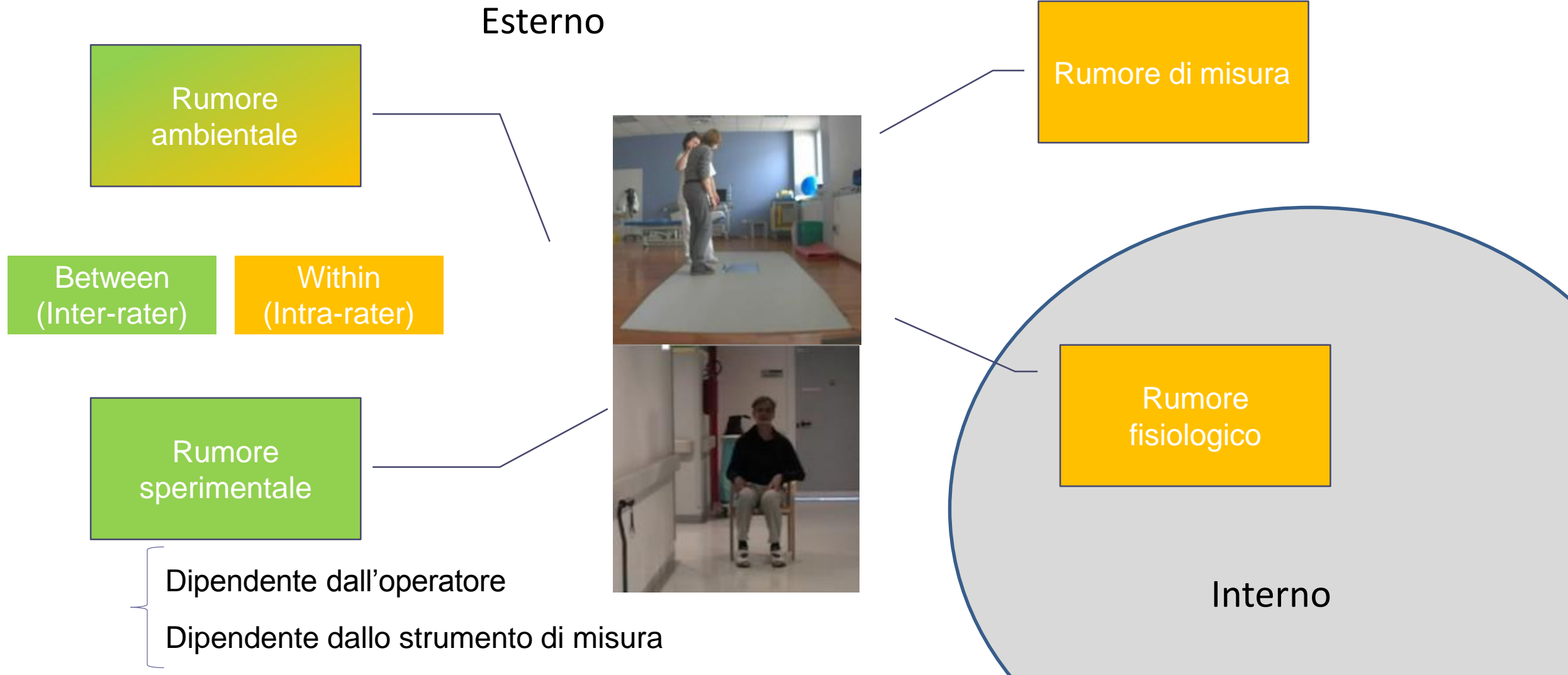
Rumore di misura

Rumore
sperimentale

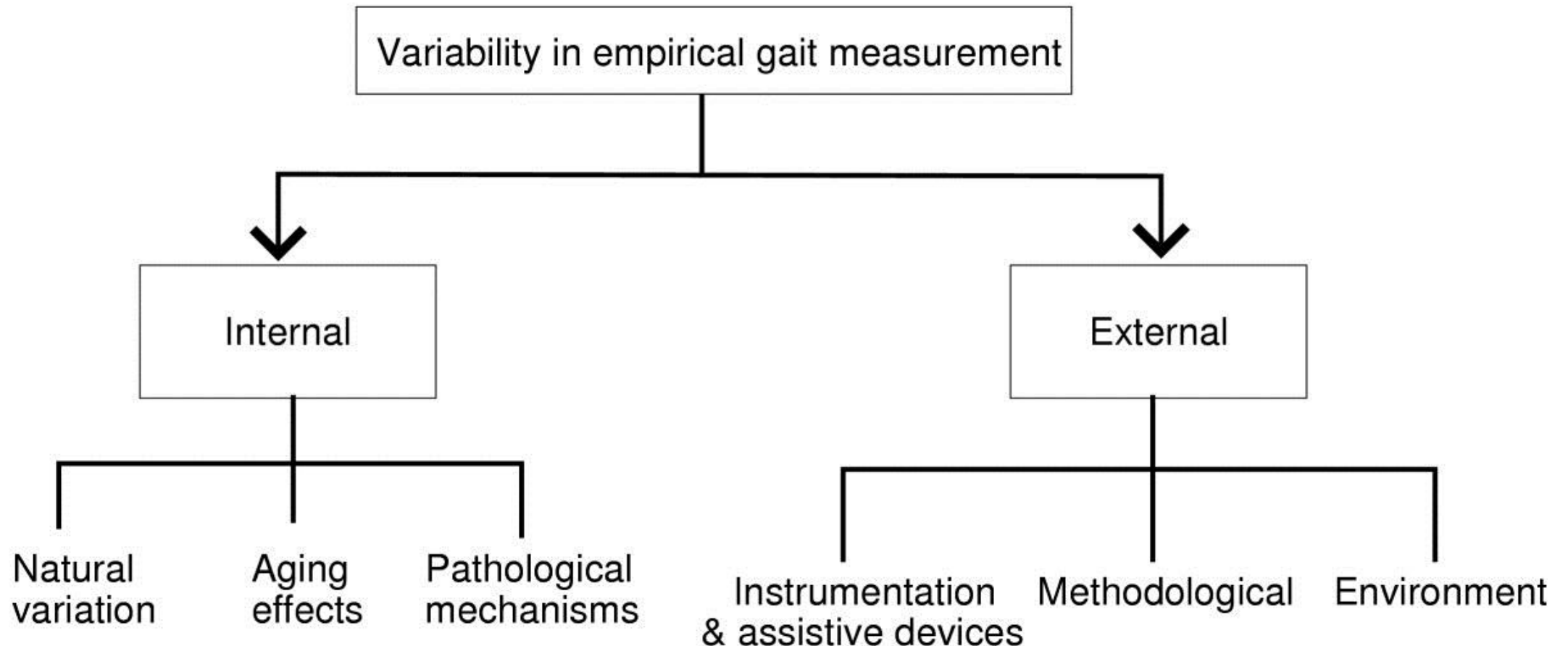
Rumore
fisiologico



Alcune sorgenti di rumore



Un esempio: sorgenti di variabilità nella misura del cammino





Valutazione della qualità di una misura

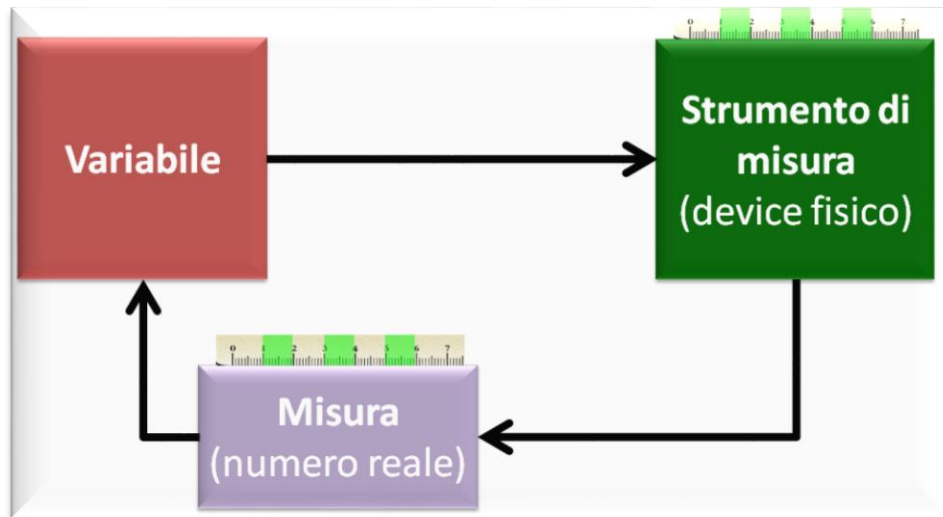
- E' impossibile controllarle tutte (o quanto meno tutte quelle che non portano l'informazione che stiamo cercando), e quindi annullarle a priori. Al più si può provare a minimizzarle. A posteriori poi si dovrà cercare di **quantificarne gli effetti** residui...
- Come controllare se e quanto le varie sorgenti di rumore (e quindi di variabilità spuria) affliggono le nostre misure strumentali nel contesto riabilitativo?



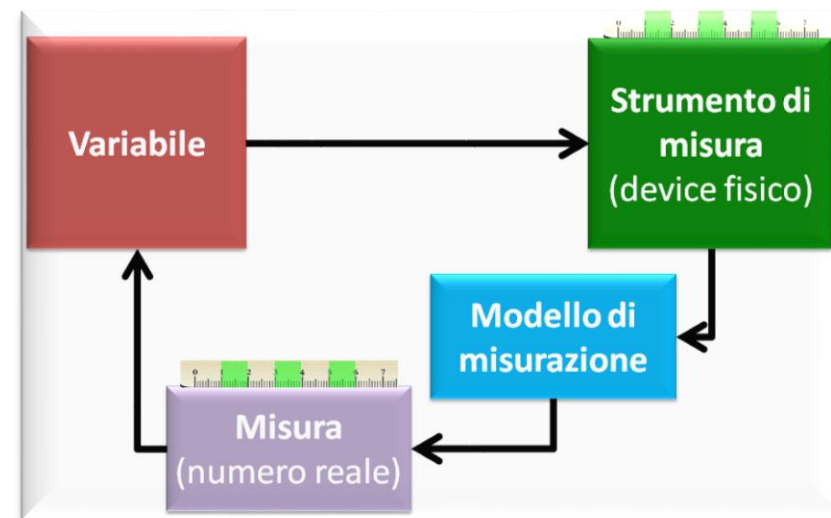
Perché si misura?

- **Per distinguere** → **Analisi cross-sezionale**
- **Per quantificare un cambiamento** → **Analisi longitudinale**
- **Per predire l'evoluzione di un sistema** → **Modelli predittivi**

Le variabili che ci interessa misurare hanno caratteristiche diverse



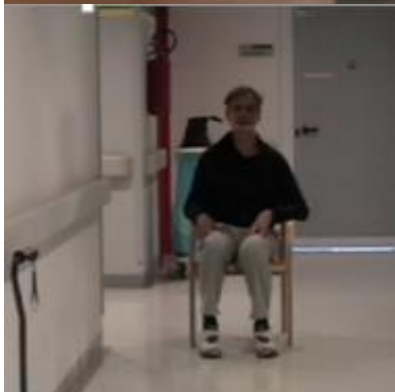
Grandezze **fondamentali** o estensive
(es. lunghezza, massa, tempo,...)



Grandezze **derivate** o intensive
(es. temperatura, velocità, ...)

Scienze fisiche
(variabili manifeste)

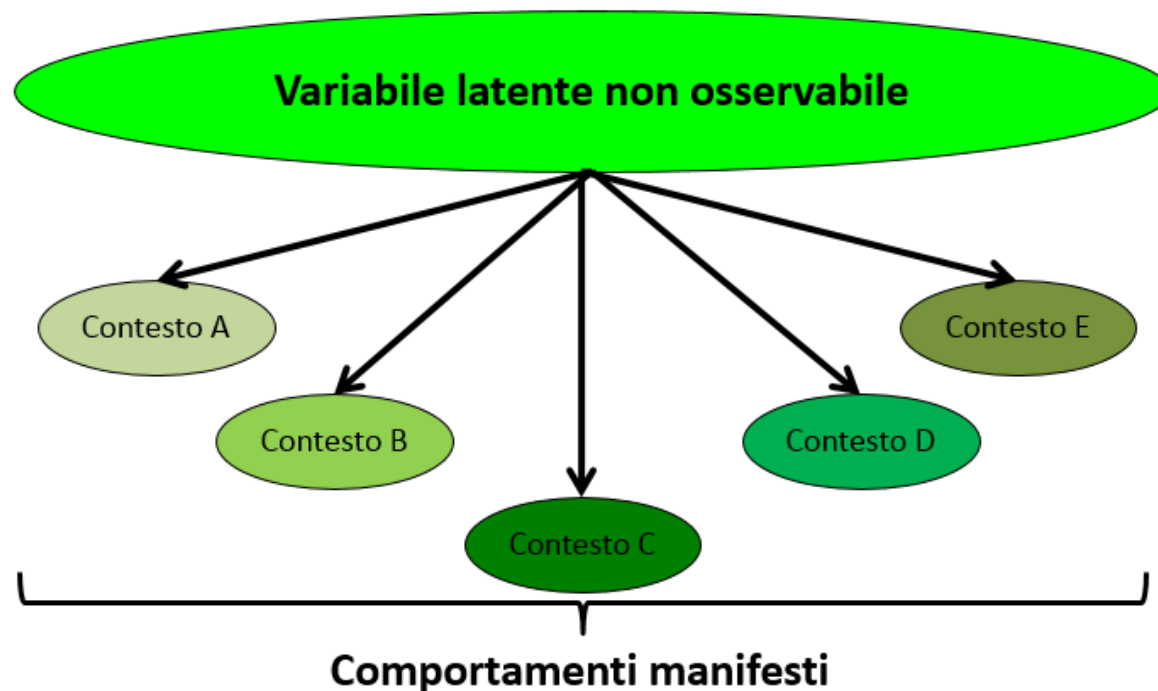
Le variabili che ci interessa misurare hanno caratteristiche diverse



(es. autonomia, qualità della vita, equilibrio, ...)

Scienze umane
(«comportamenti»)

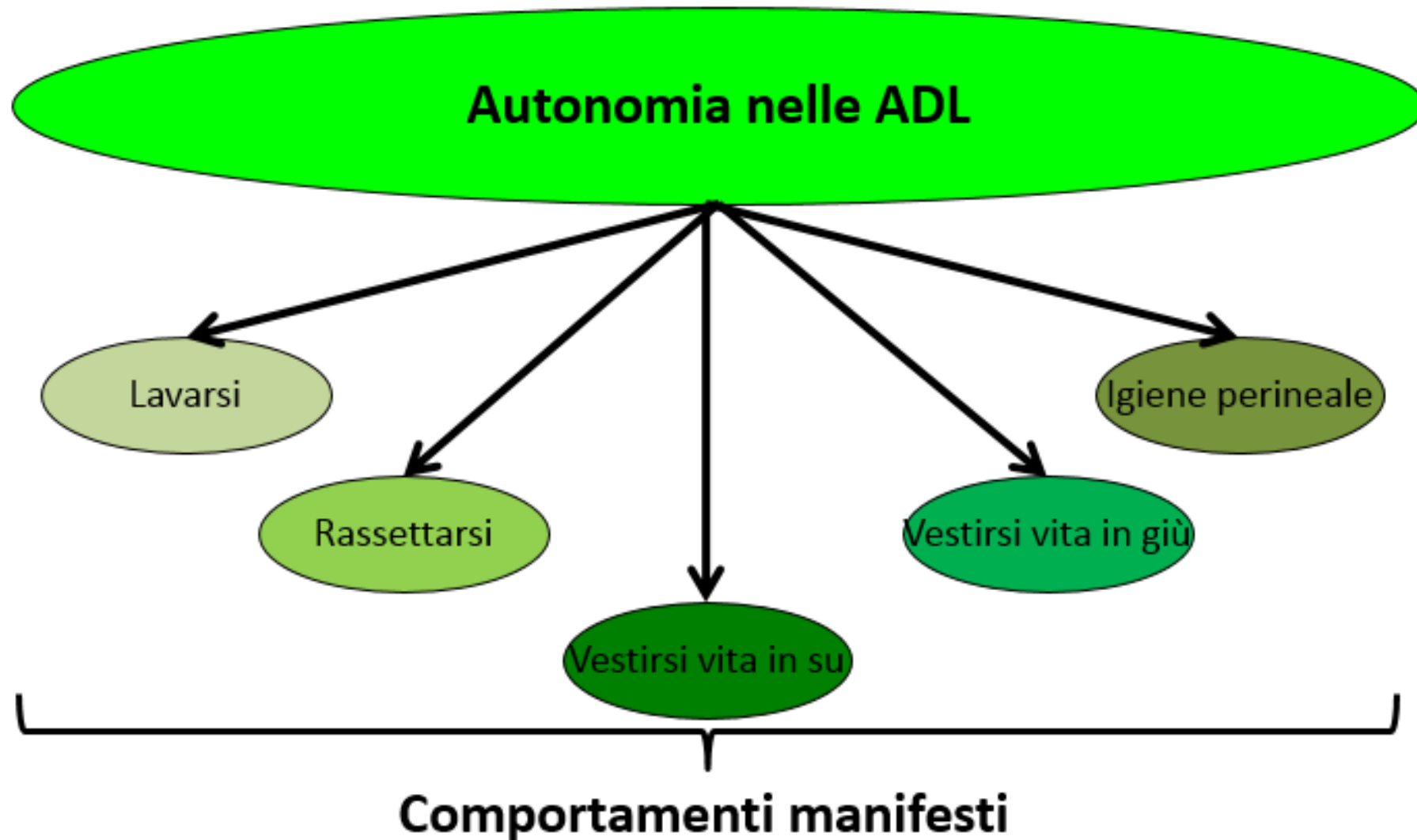
Le variabili che ci interessa misurare hanno caratteristiche diverse



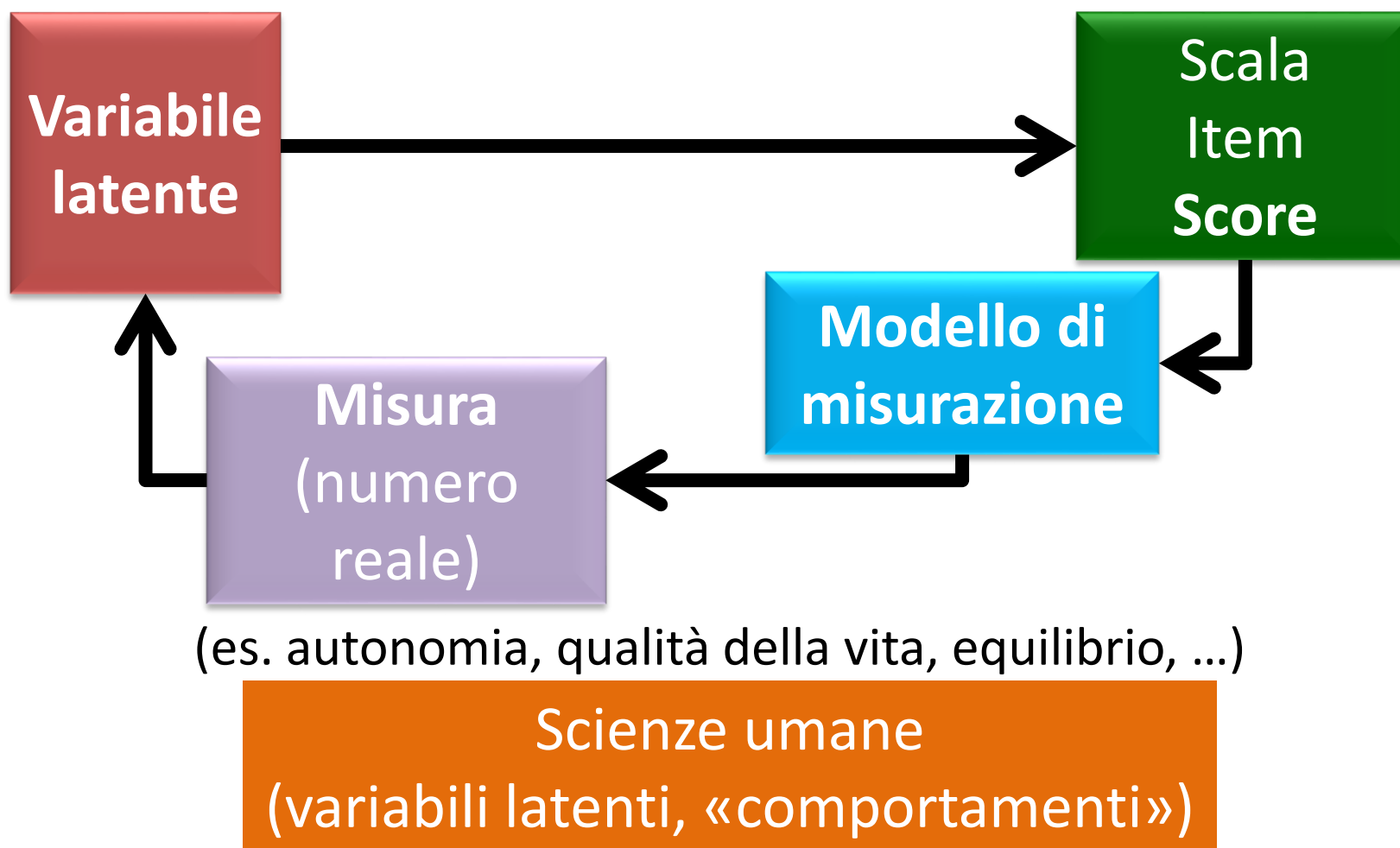
(es. autonomia, qualità della vita, equilibrio, ...)

Scienze umane
(variabili **latenti**, «comportamenti»)

● ● ● | Variabile latente: esempio



Le variabili che ci interessa misurare hanno caratteristiche diverse



Proprietà di uno strumento di misura in questo contesto

Le proprietà essenziali di una misura, ove applicata ad un problema delle scienze umane, sono:

- **Validità**
- **Attendibilità**
- **Sensibilità**
- **Trasferibilità**

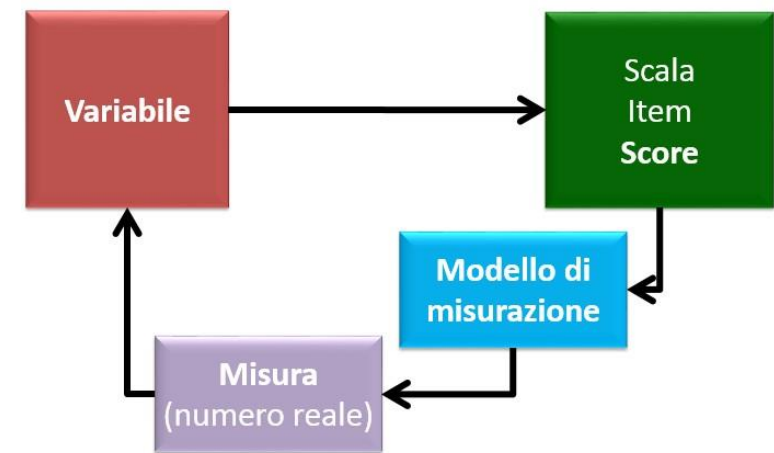


Validità

- Validità

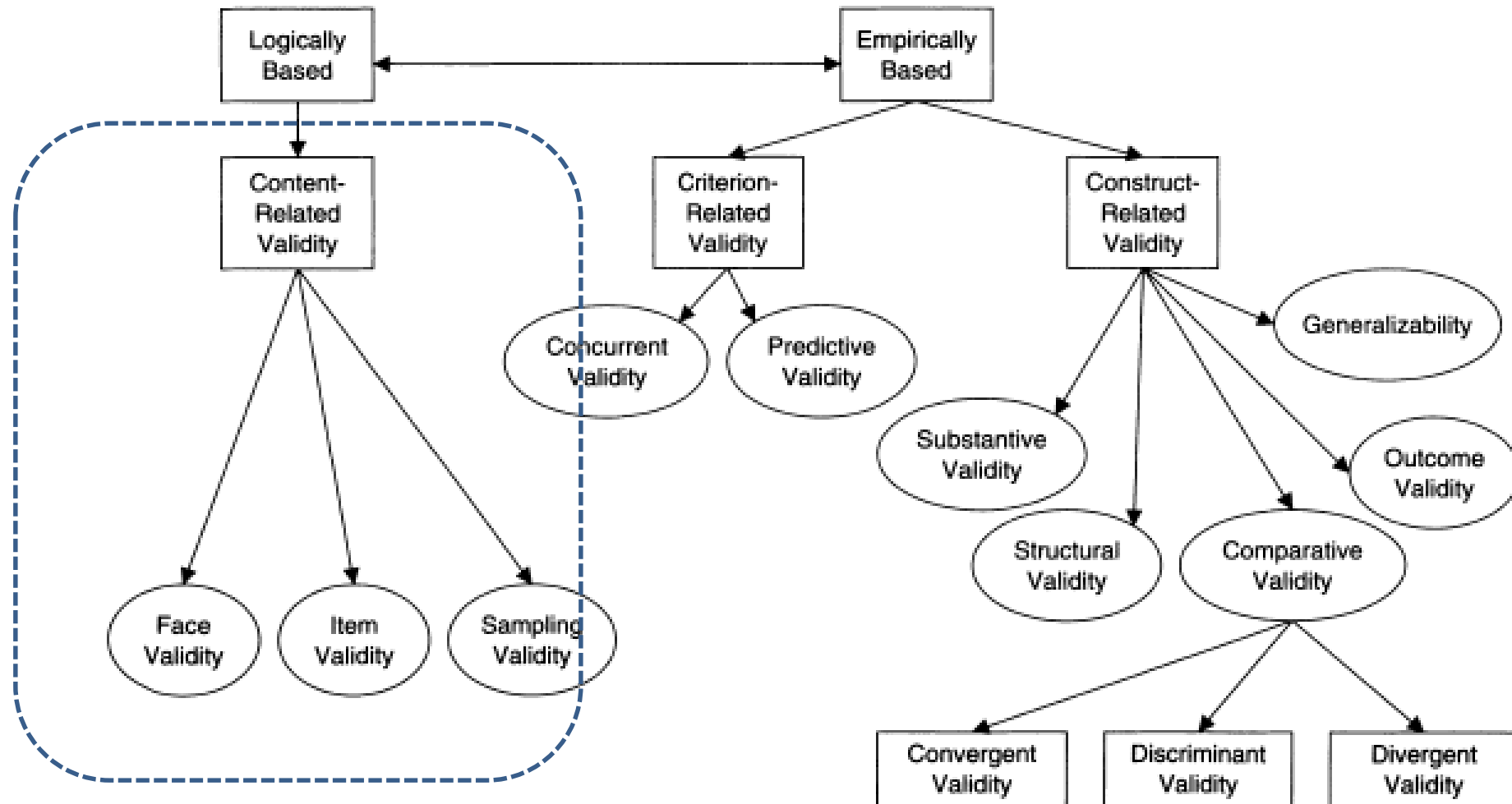
- Lo strumento misura ciò che vuole misurare

- Nel caso di sistemi che includono/forniscono più misure (es. scala a più items), un insieme di misure include aspetti effettivamente rappresentativi dell'ambito che si vuole misurare (*content validity*);
- Una misura è **coerente con il modello teorico** in base al quale è stata elaborata (*construct validity*);
- La misura, acquisita con uno strumento A, **correla** con quella acquisita con lo strumento B, già accettato come valido (*concurrent validity*);
- La misura **consente di predire** una situazione futura (*predictive validity*);
- Lo strumento di misura è effettivamente **utilizzabile** nel contesto per il quale è stato progettato (*ecologic validity*).



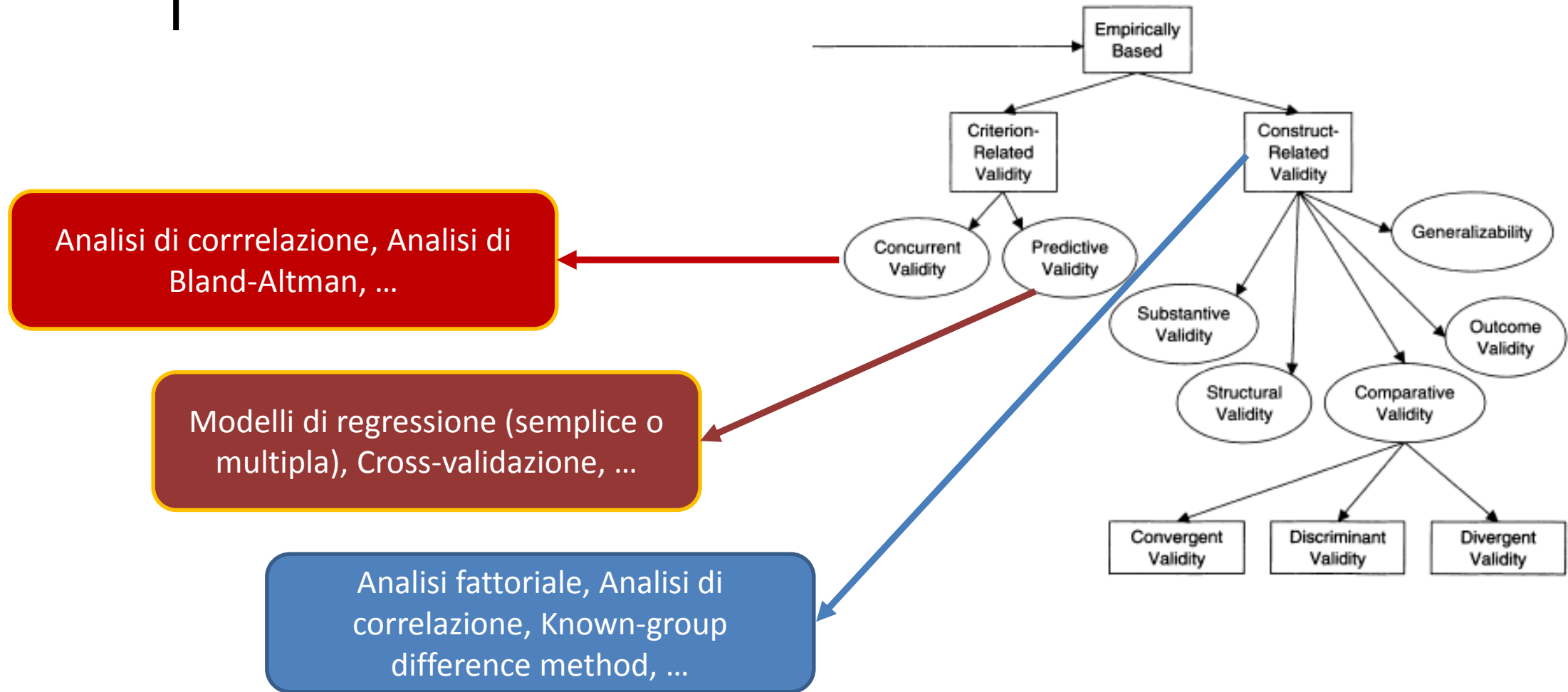
Criterion validity

Esistono vari tipi di Validità



Analisi non supportabile da alcuna evidenza statistica

Analisi di Validità: strumenti statistici





Attendibilità

Agreement e Reliability (consistenza, riproducibilità, affidabilità,...)

- **Studio valutativo:** La valutazione di uno stesso caso (stessa grandezza, misurata nello stesso soggetto nelle medesime condizioni) deve dare lo stesso esito
- **Studio discriminativo:** pazienti nelle stesse condizioni devono ricevere la stessa valutazione, indipendentemente dal valutatore, dal momento e dal contesto.
- **E' condizione necessaria ma non sufficiente per la validità:** non ci si può fidare di una misura/un test che fornisce uscite diverse su prove ripetute in sequenza o usando strumenti di misura/valutatori diversi, o che non ha capacità discriminativa, ecc., tuttavia una misura/test ripetibile può non essere valida



Agreement e Reliability

- **Accordo (*agreement*)**
 - **manca di errore di misura**
 - Parametro importante nella valutazione del **cambiamento nel tempo**
 - *Proprietà caratteristica dello strumento di misura*
- **Affidabilità (*reliability*)**
 - Quanto gli individui misurati possono essere distinti tra loro ***nonostante l'errore di misura***
 - Il coefficiente mette in relazione la variazione dovuta all'errore di misura con la variazione tra individui nella popolazione
 - Parametro importante nella **discriminazione tra individui**
 - *Proprietà caratteristica dello studio specifico*



Rispondono a domande diverse

- **Agreement:** «How good is the agreement between repeated measurements?»
- **Reliability:** «How reliable is the measurement?»
- Hanno in comune una situazione di **test-retest** (misure ripetute): in diversi momenti o condizioni sperimentali, con diversi valutatori, o con lo stesso valutatore in momenti diversi.
- La quantificazione dell'**agreement** è particolarmente rilevante per misure utilizzate con *finalità valutative* (comprese le misure di variazione di una grandezza nel tempo, dovute ad esempio ad un intervento riabilitativo)
- La quantificazione della **reliability** è particolarmente rilevante per misure utilizzate con *finalità discriminative*

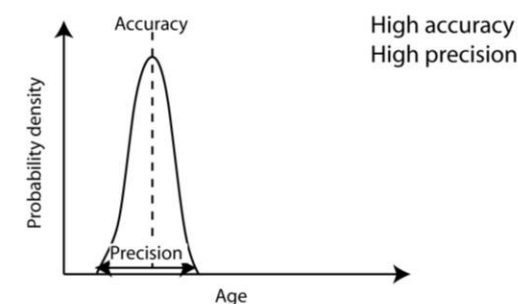
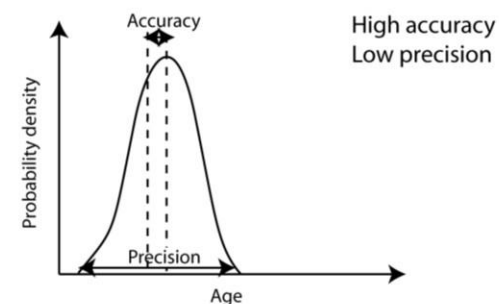
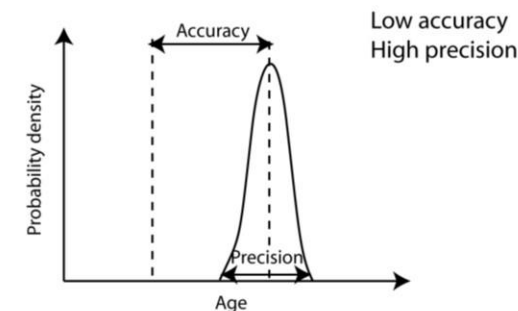
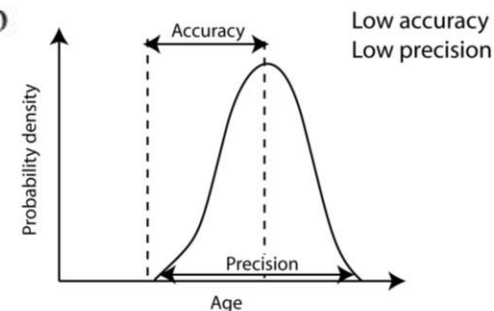
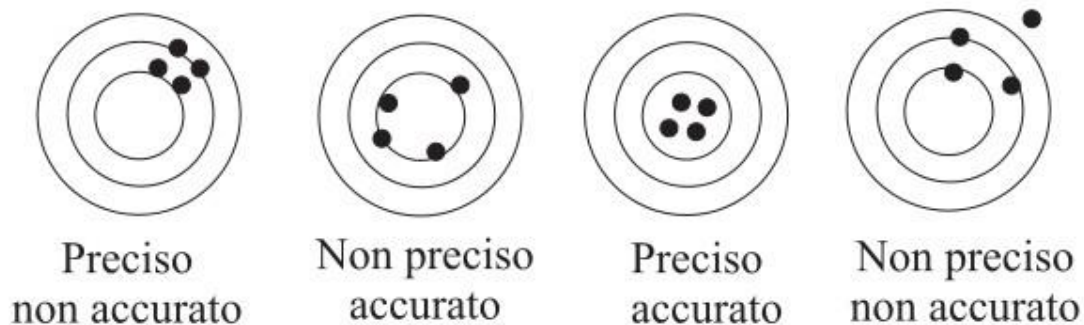


Agreement vs Reliability

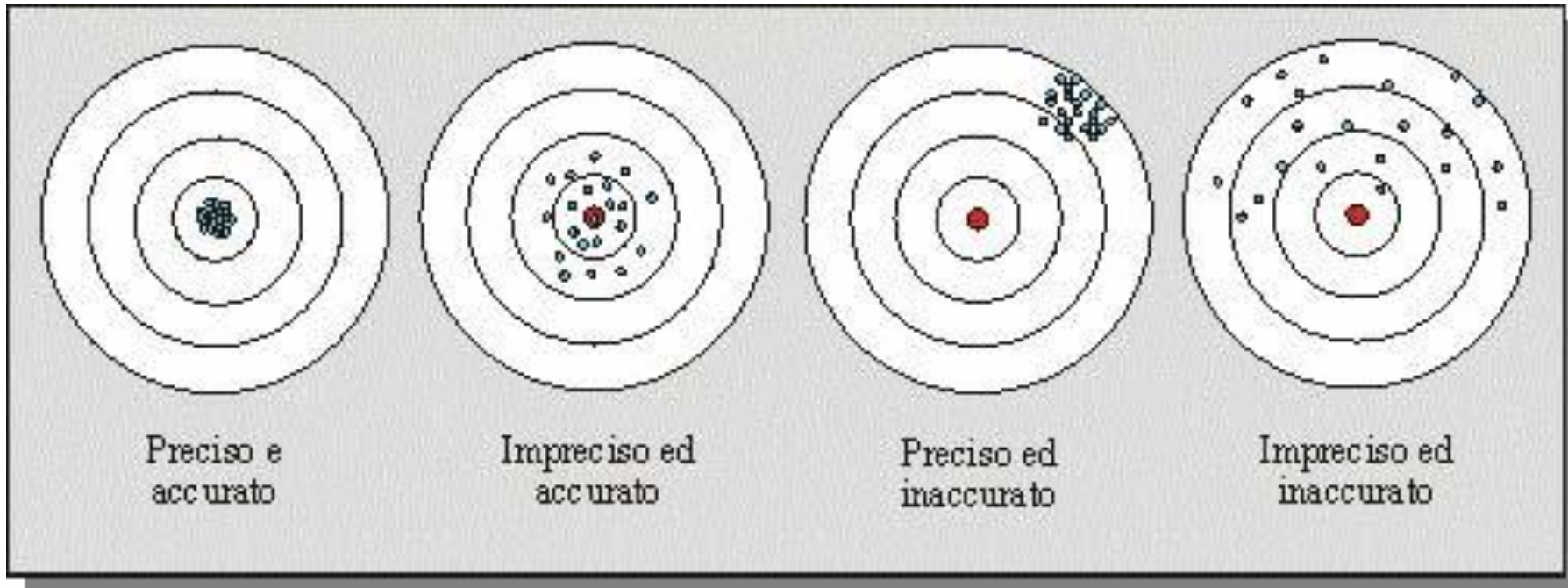
*"Discriminative instruments require a high level of reliability: that is, **the measurement error should be small in comparison to the variability between the persons that the instrument needs to distinguish.** Thus, if the differences between persons are large, a certain amount of measurement error is acceptable. For an **evaluative measurement instrument** **the variability between persons in the population sample does not matter at all; only the measurement error is important.** This measurement error should be smaller than the improvements or deteriorations that one wants to detect. If the measurement error is large, then small changes cannot be distinguished from measurement error. **The smaller the measurement error, the smaller the changes that can be detected beyond measurement error.**"*

(de Vet et al., 2006)

Proprietà di uno strumento di misura: Precisione e accuratezza



Precisione e accuratezza vs Validità e affidabilità



Misurazioni valide

Precisa
(affidabile)

Imprecisa
(non affidabile)

Misurazioni NON valide

Precisa
(affidabile)

Imprecisa
(non affidabile)

Esempio

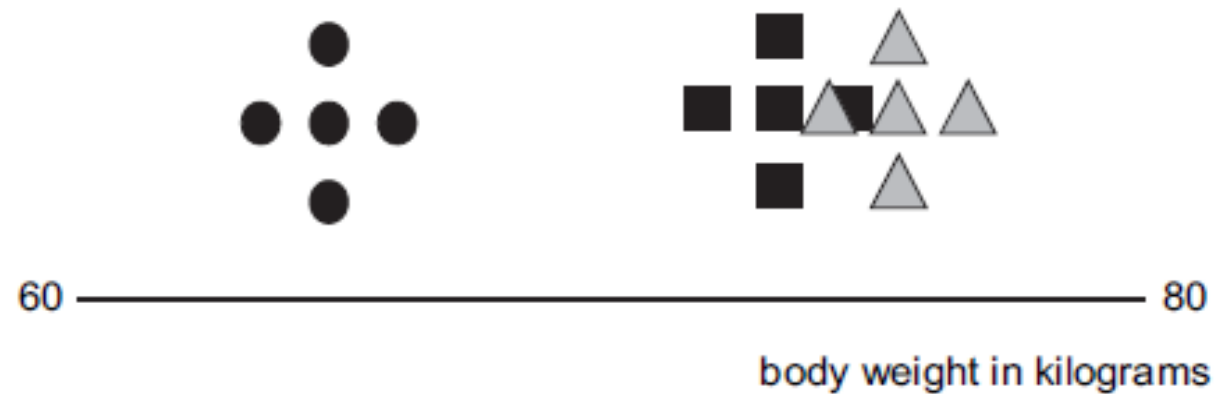


Fig. 1. Five repeated measurements of the body weights of three persons (●, ■, and ▲).

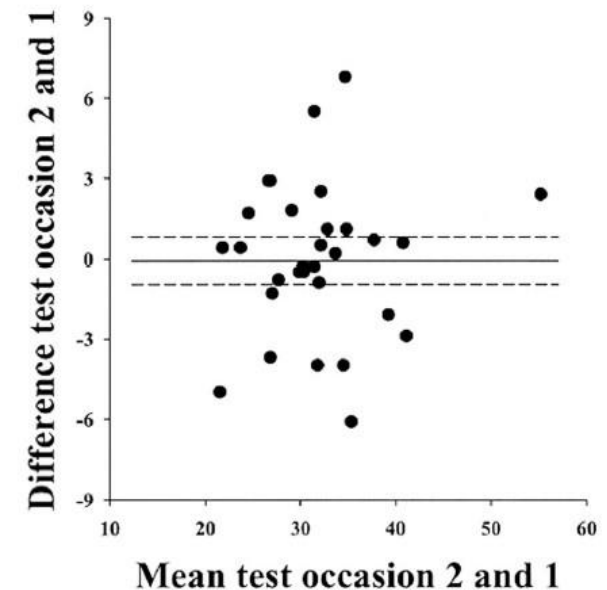
- Peso di 3 persone misurato in 5 giornate consecutive
- Le 5 misure mostrano la stessa variabilità (es. SD) nei 3 soggetti → Agreement
- Tale variabilità ci consente di distinguere bene i soggetti ● e ■ (alta affidabilità), ma non altrettanto ■ e ▲ (bassa affidabilità) → Reliability

Misure di agreement 1/3

- Bland-Altman plot e relativi Limits of Agreement (LOA)
 1. Calcolare le medie e le differenze tra le due letture corrispondenti: $(A+B)/2$ e $(A-B)$
 2. Graficare le medie in ascissa e le differenze in ordinata in uno scatterplot
 3. Calcolare media e SD delle differenze: M_{diff} , SD_{diff}
 4. Calcolare i limits of agreement: **LOA = $M_{diff} \pm 2 SD_{diff}$**

Indicatori di agreement:

- $M_{diff} \approx 0$
- $SD_{diff} \ll$



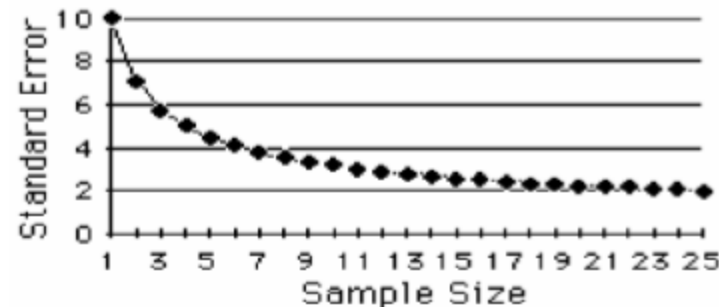
Misure di agreement 2/3

Per misure continue (come quelle fornite dai dispositivi di misura) la misura più comune per quantificare l'agreement è la **SEM** (*errore standard della media*):

$$SEM = \frac{SD_{diff}}{\sqrt{N}}$$

dove N è il numero di osservazioni per campione (quindi N=2 se andiamo a confrontare tra loro gli esiti di misurazioni effettuate 2 volte come nel caso di Bland-Altman).

Il SEM è dunque inversamente proporzionale alla radice quadrata del numero di ripetizioni e cala all'aumentare di queste.





Misure di agreement 3/3

- E' possibile anche ricavare il SEM dalla tabella dell'ANOVA

- $SEM = \sqrt{\sigma_{error}^2}$

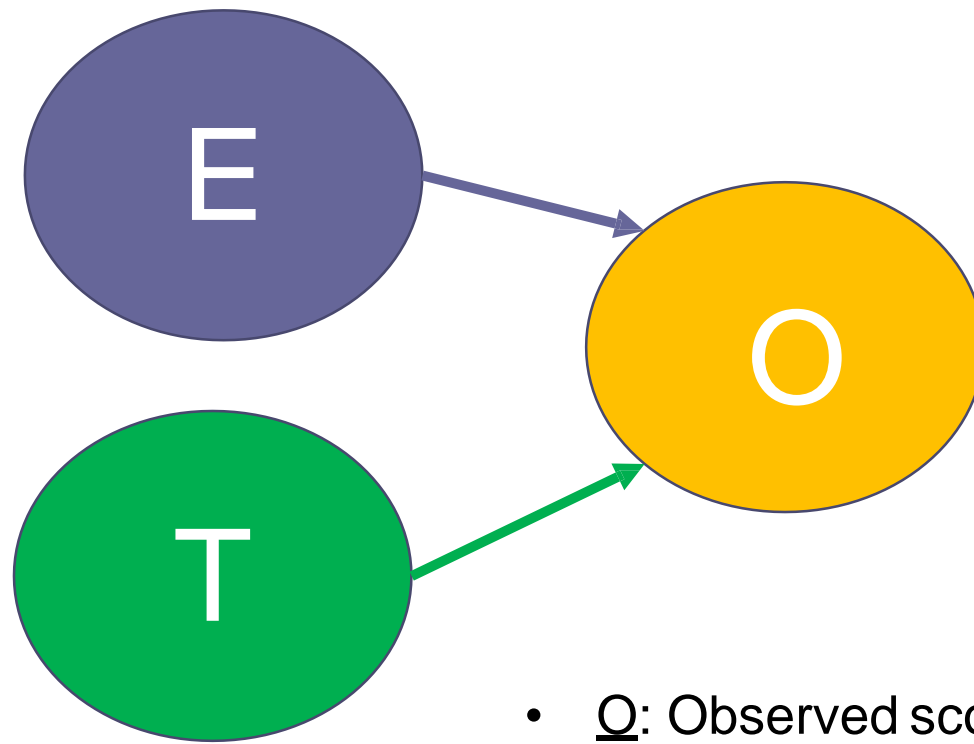
- La varianza totale osservata è costituita da tre termini:

$$\sigma^2 : \quad \sigma_{soggetti}^2, \sigma_{raters}^2, \sigma_{residua}^2$$

L'ultimo termine rappresenta l'interazione tra soggetti e raters
e σ_{error}^2 rappresenta il termine relativo all'errore quadratico medio
ottenuto dall'ANOVA :

$$\sigma_{error}^2 = \sigma_{residua}^2$$

Modello classico del processo di misura



- O: Observed score
- T: True score
- E: Error score

$$O = T + E$$

$$O - E = T$$



Misure di reliability (teoria classica)

$$Reliability = \frac{var(T)}{var(O)} \approx \frac{var(O) - var(E)}{var(O)}$$

$$O = T + E$$

$$T = O - E$$

Auspicio: $E \rightarrow 0 \Rightarrow O \approx T \Rightarrow reliability \approx 1$

E' un indice di quanto la varianza della misura è libera dalla varianza d'errore



Misure di reliability (teoria classica)

$$Reliability = \frac{var(T)}{var(O)} \approx \frac{var(O) - var(E)}{var(O)}$$

Più in generale:

$$Reliability = \frac{variabilità\ tra\ soggetti}{variabilità\ tra\ soggetti + \text{errore di misura}}$$



Analisi di Reliability: strumenti

Vogliamo quantificare e confrontare tra loro i possibili contributi alla **varianza** delle misure effettuate

La varianza delle misure viene suddivisa in varianza dovuta all'appartenenza ad una particolare categoria dell'unica variabile indipendente (varianza *between*), e ad altre cause, errore compreso (varianza *within*).

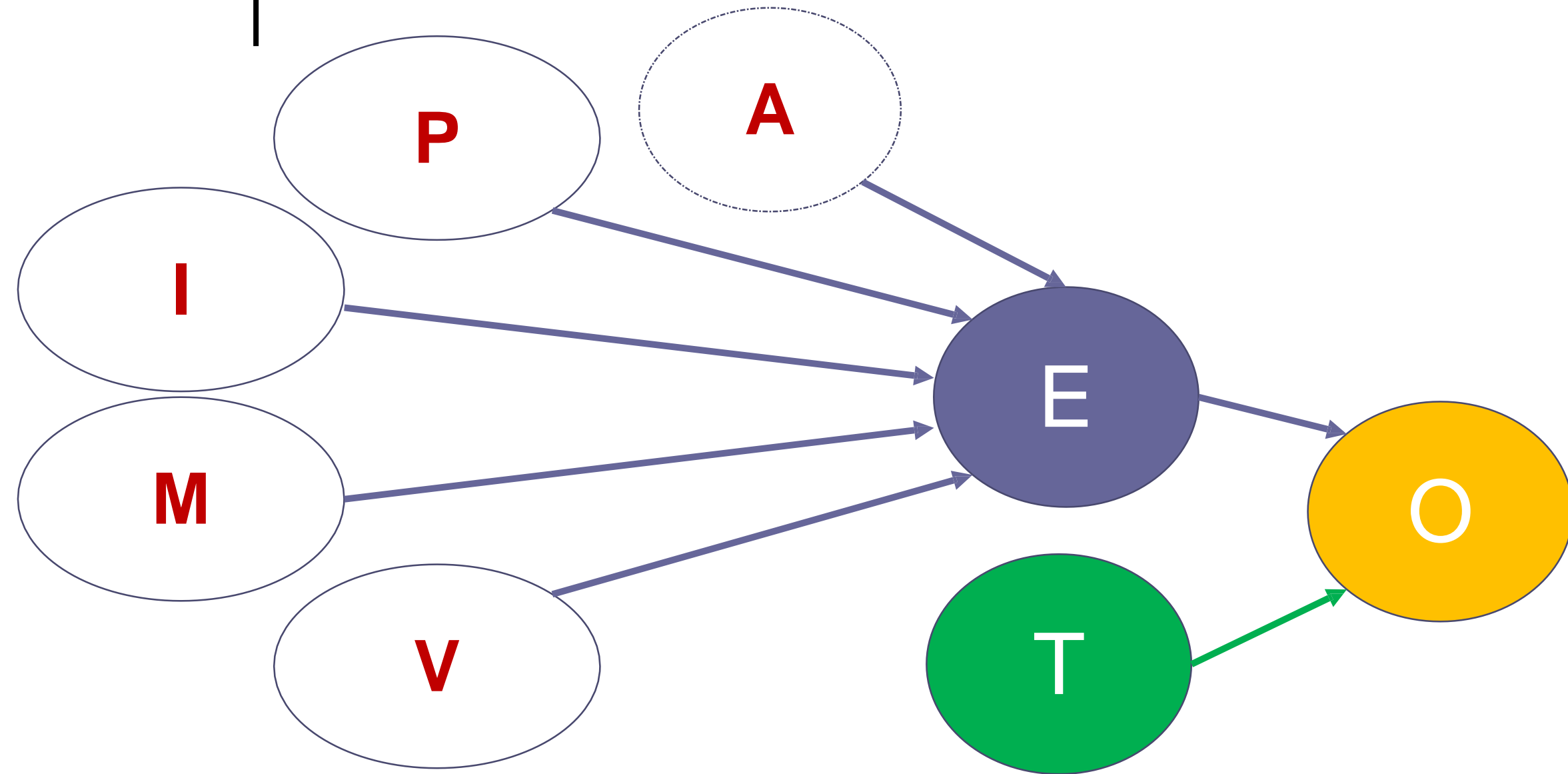
La varianza *between* è quella dovuta alle differenze delle misure fra le unità di analisi (soggetti), mentre la varianza *within* è quella dovuta all'errore casuale di misurazione, indipendentemente dalle variazioni delle misure vere.



ANOVA a una via



Le principali sorgenti di rumore





Le principali sorgenti di rumore

Partecipante (misurando): umore, motivazioni, affaticamento, memoria, condizione di salute, fluttuazioni, pratica precedente, conoscenza specifica e familiarità con il test,...

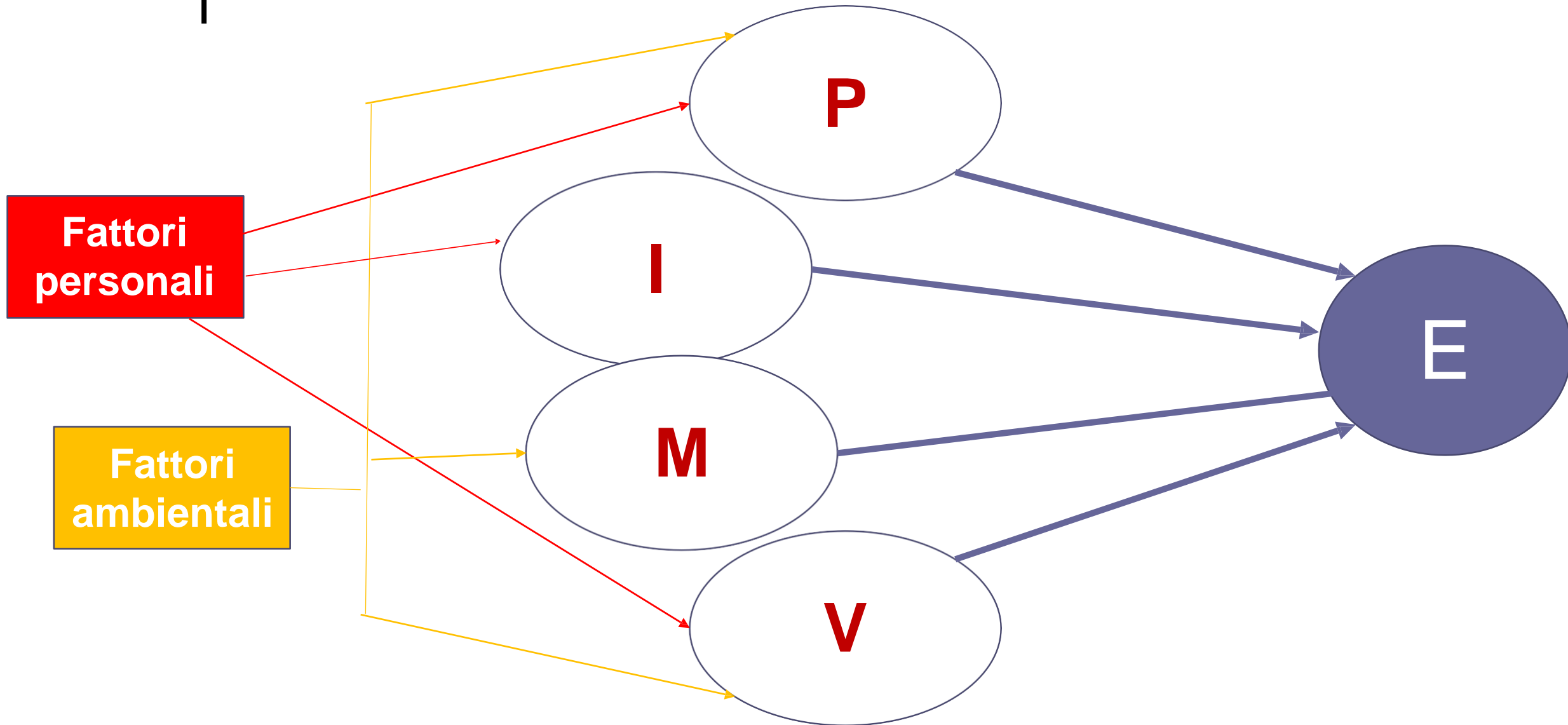
Istruzioni e test (protocollo/procedura): chiarezza e completezza delle istruzioni, aderenza alle istruzioni, consistenza nel fornire le istruzioni da parte del valutatore, ...

Misura (strumenti di misura): mancanza di calibrazione, rumore elettronico, meccanico, termico, errori nel modello di misurazione o inadeguatezza dello strumento prescelto in relazione al test, al valutatore e al partecipante,...

Valutatore (rater): competenza, esperienza e concentrazione del valutatore in relazione alla complessità della valutazione, familiarità con il test da valutare

Altro: altre possibili sorgenti di rumore non ricomprese nelle precedenti

● ● ● | Le principali sorgenti di rumore





Misure di reliability

- Per definizione la *reliability* risulta compresa nell'intervallo $[0,1]$.
- Viene misurata attraverso la correlazione tra diverse istanze della stessa variabile (è quindi una statistica univariata) che produce un cosiddetto **coefficiente di correlazione intra-classe** (ICC)
- Ci sono diverse formule per l'ICC che possono dare risultati diversi se applicate ai medesimi dati. Ogni formula è appropriata per situazioni specifiche che sono definite dal disegno sperimentale e dall'uso potenziale che si intende fare dei risultati.
- Praticamente, **tutti i valori dell'ICC sono calcolabili a partire da un'analisi di varianza (ANOVA) effettuata sulla matrice dei dati.**



Un modello più generale

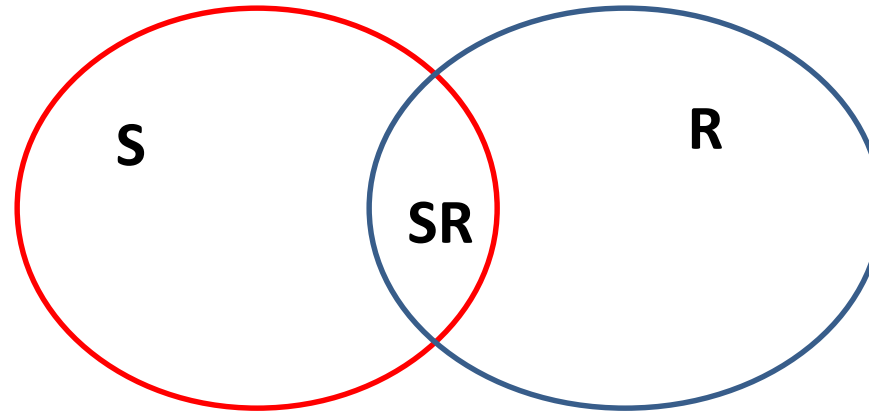
- La **teoria della generalizzabilità** (TG, o Generalizability Theory, Cronbach et al., 1972) offre un'ampia cornice di riferimento per esaminare in modo più approfondito l'attendibilità delle misure.
- La TG estende il concetto classico di attendibilità della Teoria classica dei test (TCT) cercando di identificare e quantificare il contributo di **più di una fonte di errore di misurazione**.
- L'errore di misurazione, come noto, può derivare da una molteplicità di fonti, per cui la stima della varianza di errore e quindi dell'attendibilità del test variano a seconda del modo in cui i dati vengono raccolti.



Un modello più generale

- La TG cerca di andare al di là della semplice indagine sull'accuratezza del punteggio osservato rispetto al punteggio vero, in quanto si chiede a quale livello di accuratezza il punteggio osservato permette di generalizzare il comportamento delle persone ad un universo di situazioni.
- Dal punto di vista statistico, la TG si basa sull'Analisi della varianza (ANOVA), in quanto **considera le possibili fonti di errore di misurazione come fattori** (o variabili indipendenti) in un modello di analisi della varianza fattoriale.
- L'ANOVA fattoriale consente di introdurre molteplici cause nel modello di spiegazione della varianza del punteggio osservato, che non solo contribuiscono a tale varianza indipendentemente le une dalle altre (i cosiddetti effetti principali) ma anche interagendo fra loro, da cui gli effetti di interazione.
- La TG parte dal presupposto che le cause di errore di misurazione siano multiple, per cui utilizza una cornice di riferimento statistica coerente.

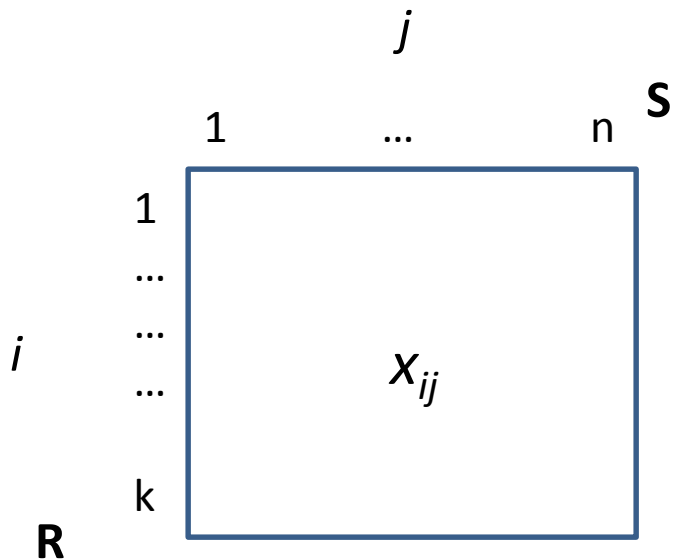
Un modello più generale



Es:

S – soggetti

R - ripetizioni



$$x_{ij} = \mu + a_i + b_j + (ab)_{ij} + e_{ij}$$

Varie tipologie di modelli e relative ipotesi

Table 1

*Analysis of Variance Models Used in Developing Intraclass
Correlation Coefficient Definitions*

Case label	Model	Assumptions
Case 1: One-way random effects	$x_{ij} = \mu + r_i + w_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$.	μ (the population mean for all observations) is constant; r_i (the row effects) are random, independent, and normally distributed with mean 0 and variance σ_r^2 ; and w_{ij} (residual effects) are random, independent, and normally distributed with mean 0 and variance σ_w^2 . Moreover, the effects r_i and w_{ij} are pairwise independent.
Case 2: Two-way random effects, with interaction	$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$.	μ and r_i are as before; c_j (the column effects) are random, independent, and normally distributed with mean 0 and variance σ_c^2 ; rc_{ij} (the interaction effects) are random, independent, and normally distributed with mean 0 and σ_{rc}^2 ; and e_{ij} (residual effects) are random, independent, and normally distributed with mean 0 and variance σ_e^2 . Moreover, all the effects are pairwise independent.

McGraw and Wong, 1996

Case 2A: Two-way random effects, interaction absent	$x_{ij} = \mu + r_i + c_j + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$.	Same as for Case 2 except that there is no interaction effect.
Case 3: Two-way mixed effect model, with interaction	$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$.	Same as for Case 2 except that c_j are fixed so that $\sum c_j = 0$, $\sum_{j=1}^k rc_{ij} = 0$, and the parameter corresponding to σ_c^2 in Case 2 is $\theta_c^2 = \sum c_j^2 / (k - 1)$.
Case 3A: Two-way mixed model, interaction absent	$x_{ij} = \mu + r_i + c_j + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$.	Same as for Case 3 except that there is no interaction effect.

Varie tipologie di modelli

Table 3
Mean Square Expectations for Analysis of Variance Models Given in Table 1

Model and source of variation	<i>df</i>	<i>MS</i>	<i>EMS</i>
Case 1: One-way random effects model			
Between rows	$n - 1$	MS_R	$k\sigma_r^2 + \sigma_w^2$
Within rows	$n(k - 1)$	MS_W	σ_w^2
Case 2: Two-way random model with interaction			
Between rows	$n - 1$	MS_R	$k\sigma_r^2 + \sigma_c^2 + \sigma_e^2$
Within rows	$n(k - 1)$	MS_W	$\sigma_r^2 + \sigma_c^2 + \sigma_e^2$
Between columns	$k - 1$	MS_C	$n\sigma_c^2 + \sigma_r^2 + \sigma_e^2$
Error	$(n - 1)(k - 1)$	MS_E	$\sigma_r^2 + \sigma_c^2$
Case 2A: Two-way random model, interaction absent			
Between rows	$n - 1$	MS_R	$k\sigma_r^2 + \sigma_e^2$
Within rows	$n(k - 1)$	MS_W	$\sigma_r^2 + \sigma_e^2$
Between columns	$k - 1$	MS_C	$n\sigma_c^2 + \sigma_e^2$
Error	$(n - 1)(k - 1)$	MS_E	σ_e^2
Case 3: Two-way mixed model with interaction			
Between rows	$n - 1$	MS_R	$k\sigma_r^2 + \sigma_e^2$
Within rows	$n(k - 1)$	MS_W	$\theta_c^2 + \frac{k}{k - 1}\sigma_c^2 + \sigma_e^2$
Between columns	$k - 1$	MS_C	$n\theta_c^2 + \frac{k}{k - 1}\sigma_c^2 + \sigma_e^2$
Error	$(n - 1)(k - 1)$	MS_E	$\frac{k}{k - 1}\sigma_c^2 + \sigma_e^2$
Case 3A: Two-way mixed model, interaction absent			
Between rows	$n - 1$	MS_R	$k\sigma_r^2 + \sigma_e^2$
Within rows	$n(k - 1)$	MS_W	$\theta_c^2 + \sigma_e^2$
Between columns	$k - 1$	MS_C	$n\theta_c^2 + \sigma_e^2$
Error	$(n - 1)(k - 1)$	MS_E	σ_e^2

Note. $E(MS)$ = expected mean squares; MS_R = mean square for rows; MS_W = mean square for residual sources of variance; MS_C = mean square for columns; MS_E = mean square error.

McGraw and Wong, 1996

Varie tipologie di modelli: formule per l'ICC

McGraw and Wong, 1996

Table 4
Single Score Intraclass Correlation Coefficients (ICCs) for One-Way and Two-Way Models

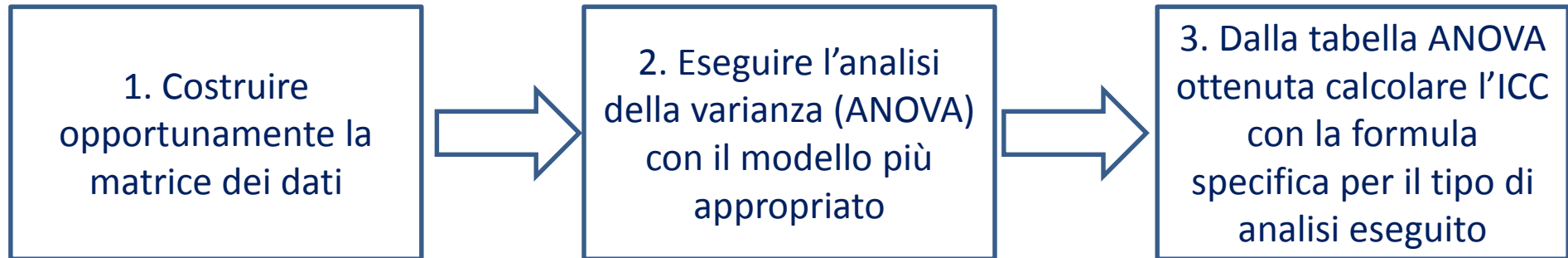
Definitions of ICCs ρ	Formulas for calculating $\hat{\rho}$	Designation	Interpretation of ICC
Row effects random			
One-way model Case 1 model $\frac{\sigma_i^2}{\sigma_i^2 + \sigma_w^2}$	$\frac{MS_R - MS_W}{MS_R + (k-1)MS_W}$	ICC(1)	The degree of absolute agreement among measurements made on randomly selected objects. It estimates the correlation of any two measurements.
Column and row effects random (two-way random effects model)			
Two-way models ^a Case 2 model $\frac{\sigma_i^2}{\sigma_i^2 + (\sigma_c^2 + \sigma_e^2)}$ or Case 2A model $\frac{\sigma_i^2}{\sigma_i^2 + \sigma_c^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$	ICC(C,1)	The degree of consistency among measurements. Also known as norm-referenced reliability and as Winer's adjustment for anchor points (Winer, 1971). In generalizability theory, this ICC estimates the squared correlation of individual measurements and universe scores.
Case 2 model $\frac{\sigma_i^2}{\sigma_i^2 + \sigma_c^2 + (\sigma_c^2 + \sigma_e^2)}$ or Case 2A model $\frac{\sigma_i^2}{\sigma_i^2 + \sigma_c^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$	ICC(A,1)	The degree of absolute agreement among measurements. Also known as criterion-referenced reliability. Estimates the Type 1 ICC for one-way, unmatched data (Rajaratnam, 1960).

Column effects fixed, row effects random (two-way mixed effect model)			
Case 3 model $\frac{\sigma_i^2 - \sigma_c^2/(k-1)}{\sigma_i^2 + (\sigma_c^2 + \sigma_e^2)}$ or Case 3A model $\frac{\sigma_i^2}{\sigma_i^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$	ICC(C,1)	The degree of consistency among measurements made under the fixed levels of the column factor. This ICC estimates the correlation of any two measurements, but when interaction is present, it underestimates reliability.
Case 3 model $\frac{\sigma_i^2 - \sigma_c^2/(k-1)}{\sigma_i^2 + \theta_c^2 + (\sigma_c^2 + \sigma_e^2)}$ or Case 3A model $\frac{\sigma_i^2}{\sigma_i^2 + \theta_c^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$	ICC(A,1)	The absolute agreement of measurements made under the fixed levels of the column factor.

Note. MS_R = mean square for rows; MS_W = mean square for residual sources of variance; MS_E = mean square error; MS_C = mean square for columns.

^a In the event of data with a two-way classification for which the column variance is zero (i.e., $\sigma_c^2 = 0$ or $\theta_c^2 = 0$, depending on the model), a one-way model should be used. Thus even though test scores on k parallel tests can be classified by test and test taker, the column variance by definition is zero, which means that a one-way model applies.

Workflow generale per il calcolo dell'ICC



Intra-rater reliability

- Numero osservatori (raters): **1**
- Numero di misure ripetute dal medesimo osservatore: **k**
- Numero di soggetti analizzati: **n**

Procedura:

- ANOVA a 1 via (random effects model)
- Calcolo dell'ICC con la formula che segue

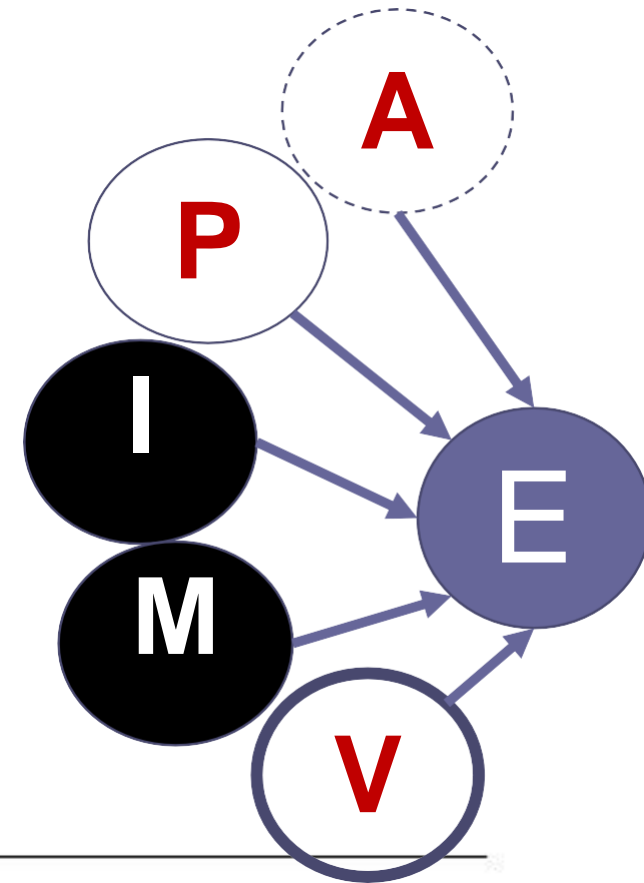


Table 3 ANOVA table of results for intra-rater reliability – one-way random effects model

						For data in Table 1	
Source of variation	Degrees of freedom (df)	Mean square (MS)	Expected mean square E(MS)	df	Between-scans day 1	$ICC (1,1) = \frac{BMS - WMS}{BMS + (k-1) WMS}$	
Between-subjects	$n-1$	BMS	$\sigma_e^2 + k\sigma_s^2$	9	13.57		
Within-subjects	$n(k-1)$	WMS	σ_e^2	10	0.41	11.55	12
						0.45	0.98

k is the number of times; **n** is the number of subjects; σ^2 = variance.

Inter-rater reliability

- Numero osservatori (raters): k (*gli unici raters di interesse*)
- Numero di misure ripetute dal medesimo osservatore: 1
- Numero di soggetti analizzati: n

Procedura:

- ANOVA a 2 vie (mixed model, raters fixed)
- Calcolo dell'ICC con la formula che segue

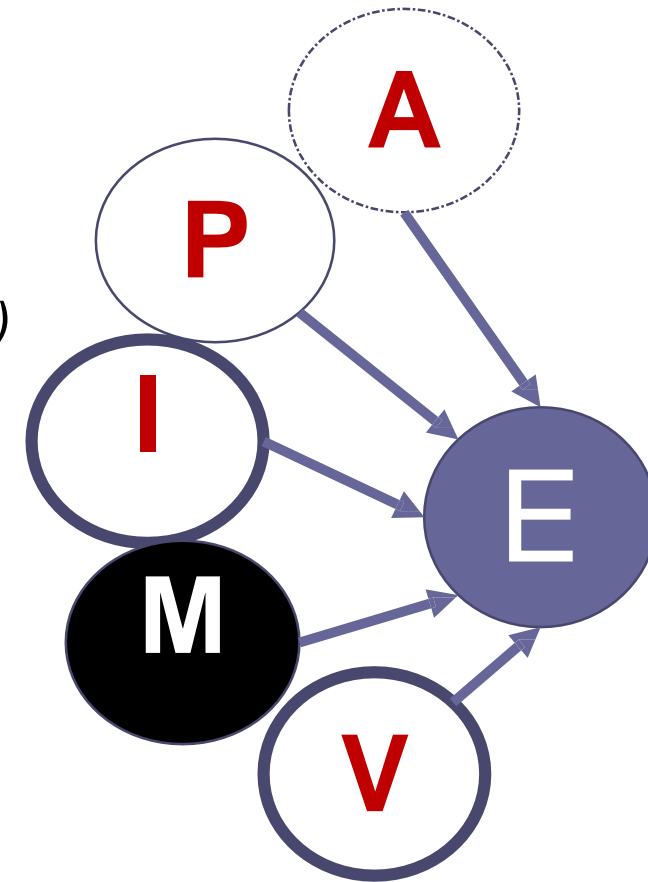


Table 2 ANOVA table of results for inter-rater reliability – two-way mixed model, raters fixed

Source of variation	Degrees of freedom (df)	Mean square (MS)	Expected mean square E(MS)
Between-subjects	$n-1$	BMS	$\sigma_e^2 + k\sigma_s^2$
Between-raters	$k-1$	RMS	$\sigma_e^2 + n/k-1 \sum \rho_j^2$
Error	$(n-1)(k-1)$	EMS	σ_e^2

$$ICC(3,1) = \frac{BMS - EMS}{BMS + (k-1) EMS}$$

k is the number of raters; n is the number of subjects; σ^2 variance. For expected mean square equations see Fleiss.¹¹

In questo caso l'analisi misura la concordanza tra gli osservatori, che sono tutti gli osservatori di interesse. Il risultato non è generalizzabile ad altri osservatori.



Inter-rater reliability (o oggettività)

- Numero osservatori (raters): ***k*** (*estratti da più ampio set di raters al quale si vuole generalizzare il risultato ottenuto*)
- Numero di misure ripetute dal medesimo osservatore: **1**
- Numero di soggetti analizzati: ***n***

Procedura:

1. ANOVA a 2 vie (mixed model)
2. Calcolo dell'ICC con la formula che segue

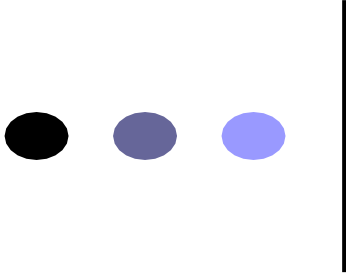
$$ICC(2,1) = \frac{BMS - EMS}{BMS + (k-1)EMS + k(RMS-EMS)/n}$$

In questo caso l'analisi misura la concordanza tra gli osservatori, ma anche la loro interscambiabilità.



Come valutare l'ICC?

- Non esistono dei valori di cut-off universali e condivisi
- Proposte di Fleiss
(1986): $ICC \geq 0,75$ - ***excellent*** reliability
 $0,4 < ICC < 0,75$ - ***fair-to-good*** reliability
 $ICC \leq 0,4$ - ***poor*** reliability



Comparison of conventional scales and wireless, wearable sensor-derived tests of mobility-related functioning

DATA	USUAL METHOD	mHEALTH SENSORS
Type of physical activity	Self-report diary or checklist; observe in lab; video; short distance timed walk or distance walked in 2-6 min.	Activity pattern-recognition algorithms; walk, cycle, leg exercises identifiable by sensor data processing
Quantity Frequency/duration	Observation; inertial movement/step counts if accelerations high enough	Directly measure wave forms of individual components and whole actions
Quality	Laboratory motion analysis or pressure mat system	Compare each leg during step cycle in context of environs
Location of activity	Self report; lab	Anywhere; global positioning & ambient context sensing for site identification
Reliability	Inter-rater; test-retest	Ground truth measurement v. sensor-based algorithm
Validity	Content/construct for each scale	Face validity; responsiveness
Statistical testing	Ordinal scales of physical functioning	Interval / ratio scale data
Data entry	Computer	Smartphone, tablet
Human factors	Train examiners in test administration	Train participants in a culture of technology
Regulation	Local Institutional Review Board and HIPAA	Local IRB & HIPAA; possibly Food and Drug Administration



Sensibilità

- *Sensitivity* ↔ *Responsiveness*
- Ne esistono 25 definizioni e 31 formule (de Vet et al., 2003)
- **Sensibilità** (Kellner, 1992)
 - Capacità di una misura (scala) di discriminare tra differenti gruppi di pazienti che presentano lo stesso disturbo (*misurazione discriminatoria*)
 - ... e di riflettere cambiamenti clinici rilevanti dovuti ai trattamenti (*misurazione valutativa*)
- Parametro essenziale per le misurazioni valutative del cambiamento lungo il tempo
 - Misure di outcome in studi sull'effetto del trattamento
- Può essere considerata come *validità longitudinale*
 - Uno strumento dovrebbe cogliere un cambiamento in una persona che effettivamente cambia, e non o non solo piccoli cambiamenti in persone che rimangono stabili nel tempo



Sensibilità

- *Distribution based, effect size*
 - Lo strumento più sensibile è quello che rileva il maggior miglioramento nella funzione/fenomeno
- *Anchor based, external standard*
 - Il miglioramento rilevato da due strumenti è correlato con uno standard esterno, ad esempio il giudizio del clinico o del paziente
 - Lo strumento più sensibile è quello che correla meglio con lo standard esterno



Misure di sensibilità

CLINICALLY IMPORTANT CHANGES

The smallest real difference (SRD) is defined by

$$\text{SRD} = 1.96 \times \text{SEM} \times \sqrt{2} \quad [10]$$

The 95% SRD is defined by

$$95\% \text{ SRD} = \bar{d} \pm \text{SRD} \quad [11]$$

The SRD% is defined by

$$\text{SRD\%} = (\text{SRD}/\text{mean}) \times 100 \quad [12]$$

where mean in Equation 12 is the mean of all the data from the two test occasions.

- Talvolta SRD è anche indicato come Smallest Detectable Change (SDC)



Trasferibilità: la relatività del concetto di validità

o Trasferibilità

- Uno strumento può essere validato in uno studio, ma non essere adatto in un'altra situazione
 - La validità è in funzione della popolazione selezionata
 - Può avere coefficienti di correlazione accettabili ma deboli con altri strumenti (validità di costrutto)
 - della misura dell'effetto indagata
 - Delle circostanze della misurazione
 - Dell'esecuzione della misurazione (ottimale o subottimale)
- È necessario valutare attentamente le circostanze dello studio di valutazione e generalizzare ad altri ambiti con cautela



Principali proprietà psicometriche

Fonte: Shirley Ryan Abilitylab

<https://www.sralab.org/rehabilitation-measures>

Standard Error of Measurement (SEM)	<p>The Standard Error of Measurement (SEM) is a reliability measure that assesses response stability. The SEM estimates the standard error in a set of repeated scores.</p> <p>Clinical Bottom Line: The SEM is the amount of error that you can consider as measurement error.</p> <p><i>In the Rehabilitation Measures Database, the SEM was frequently pulled directly from peer reviewed journal articles. However, whenever the statistics were available in the published articles, the following equation was utilized to calculate the SEM:</i></p> <p><i>SEM = Standard Deviation from the 1st test x (square root of (1-ICC))</i></p>	<p>SEM is a measurement error in the units used in the measurement.</p>
--	--	---



Principali proprietà psicometriche

Fonte: Shirley Ryan Abilitylab

<https://www.sralab.org/rehabilitation-measures>

Minimal Detectable Change (MDC)	<p>A statistical estimate of the smallest amount of change that can be detected by a measure that corresponds to a noticeable change in ability.</p> <p>Clinical Bottom Line: The MDC is the minimum amount of change in a patient's score that ensures the change isn't the result of measurement error.</p> <p><i>In the Rehabilitation Measures Database, the MDC was frequently pulled directly from peer reviewed journal articles. However, whenever the statistics were available in the published articles, the following equation was utilized to calculate the MDC:</i></p> $MDC = 1.96 \times SEM \times \text{square root of } 2$	<p>The MDC is calculated in terms of confidence of predication. For example, MDC95 is based on a 95% confidence interval, while a MDC90 is based on a 90% confidence interval. Anytime a MDC was calculated for the Rehabilitation Measures Database, the MDC95 was used.</p>
--	--	---



Principali proprietà psicometriche

Fonte: Shirley Ryan Abilitylab

<https://www.sralab.org/rehabilitation-measures>

Minimal Clinically Important Difference (MCID)	<p>MCID represents the smallest amount of change in an outcome that might be considered important by the patient or clinician.</p> <p><u>Clinical Bottom Line:</u> The MCID is a published value of change in an instrument that indicates the minimum amount of change required for your patient to feel a difference in the variable you are measuring.</p>	<p>The MCID is typically quantified in the units used in the measurement.</p>
---	--	---



Principali proprietà psicometriche

Fonte: Shirley Ryan Abilitylab

<https://www.sralab.org/rehabilitation-measures>

Test-retest Reliability	Establishes that an instrument is capable of measuring a variable with consistency.	<u>Clinical Bottom Line:</u> If you are planning to use an instrument for individual decision-making , it is recommended that you use an instrument with an ICC > 0.9 . If you are planning to use the instrument to measure progress of a large group (as in research), an instrument with an ICC > 0.7 is acceptable.
Interrater Reliability	Determines variation between two or more raters who measure the same group of subjects.	<u>Excellent Reliability:</u> ICC > 0.75 <u>Adequate Reliability:</u> ICC 0.40 to < 0.74 <u>Poor Reliability:</u> ICC < 0.40
Intrarater Reliability	Determines stability of data recorded by one individual across two or more trials.	See Interrater Reliability Criteria



Proprietà psicometriche di alcuni strumenti di valutazione

Weekly Task # 1
presentato da:

Daniela Portas, Valeria Brugattu, Manuela Lorenzetti

Falls Efficacy Scale (FES)

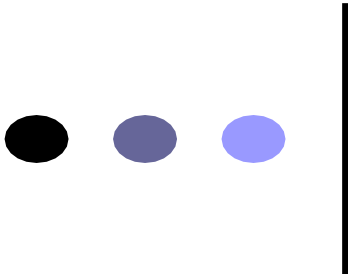


Obiettivi:

- Valutare la percezione dell'equilibrio e della stabilità durante le attività giornaliere
- Valutare la paura di cadere negli anziani
- Valutare i soggetti con diagnosi di sclerosi multipla (SM)

Dominio ICF

- Attività
- Partecipazione



Descrizione

10 item progettati per valutare il livello di confidenza che i soggetti hanno nella loro capacità di compiere dieci task giornalieri senza cadere, come indicatore di come la paura di cadere possa compromettere le performance fisiche.

Ogni item è valutato con un punteggio che va da 1 («very confident») a 10 («not confident at all»)

I punteggi vanno sommati e il punteggio totale può andare da 10, nel migliore dei casi, fino a 100, nel caso peggiore.

I punteggi più bassi indicano una maggiore fiducia da parte del soggetto nelle proprie abilità, mentre i punteggi più alti indicano la mancanza di fiducia e una maggiore paura di cadere.

Activity:	Score: 1 = very confident 10 = not confident at all
Take a bath or shower	
Reach into cabinets or closets	
Walk around the house	
Prepare meals not requiring carrying heavy or hot objects	
Get in and out of bed	
Answer the door or telephone	
Get in and out of a chair	
Getting dressed and undressed	
Personal grooming (i.e. washing your face)	
Getting on and off of the toilet	
Total Score	

Cut off

SEM

(standard error of measurement)

MDC

(minimal detectable change)

>80 aumenta il rischio di caduta

Non stabilito

Non stabilito

>70 indica la paura di cadere

(Tinetti et al, 1990; $n = 74$ community dwelling elderly persons)

Reliability

Interrater/Intrarater Reliability

**Pazienti
geriatrici**

accettabile ($r=0.71$)

Non stabilito

**Pazienti colpiti da
ictus**

eccellente ($ICC=0.97$)

(Hellstrom & Lindmark, 1999; $n = 30$; mean age = 65 (11) years; stroke onset between 5 and 84 months prior to assessment)

Criterion Validity

Pazienti geriatrici

- Eccellente se correlato con la Balance Scale (ABC) ($r=0.86$)
- Eccellente se correlato con SAFE ($r=0.67$)
- La FES non è in grado di identificare cosa limita l'attività dei pazienti
- La FES non riesce ad identificare gli individui in base alla storia di cadute

(Hotchkiss et al, 2004; $n = 118$ community dwelling individuals 60 years and older; mean age = 75.8(60-99) years)

Sclerosi multipla

- Accettabile se correlato con Timed Up and Go ($r=0.535$)
- Eccellente se correlato con il Dinamic Gait Index ($r=-0.601$)
- Eccellente se correlato con il Functional Reach ($r=-0.612$)

(Cakt et al, 2010; $n = 45$ patients with multiple sclerosis; mean age = 37.9 (10.43) years)

Morbo di Parkinson

I pazienti caduti mostrano una maggiore paura e quindi un minore grado di confidenza rispetto a quelli che non cadono.

Correlazione della Tinetti FES con le altre misurazioni		
Misurazioni	Potenza di correlazione	Correlazione
Freezing	Scarso	Tau-B=0.21, $p=0.01$
Numero di cadute	Adeguate	Rho=0.32, $p=0.001$
Uso di dispositivi di assistenza	Adeguate	Tau-B=0.35, $p=0.001$
MMSE (inverse)	Adeguate	R=-0.41, $p<0.001$

Construct Validity

Pazienti geriatrici

Scarsa correlazione con l'età $r=(-0.23)$
Eccellente correlazione con l'equilibrio $r=(0.66)$
Eccellente correlazione con l'andatura $r=(0.67)$
Eccellente correlazione con la mobilità $r=(0.71)$
Adeguate correlazione con la storia delle cadute $r=(-0.47)$
Scarsa correlazione con le condizioni mediche $r=(-0.18)$
Adeguate correlazione con lo stato di salute $r=(0.36)$

Pazienti con danni cerebrali

	Punteggio FES	Strength
Età	0.254	scarso
Berg Balance Scale	0.467	adeguato
Assistive Device	-0.399	adeguato
Dynamic Gait Index score	0.473	adeguato

Medley et al., 2006; $n = 26$, mean = 41.9 (12.4); mean time since injury = 9 years (10); community-dwelling participants; 15 subjects did not use assistive device

Pazienti con il Parkinson

Il FES non è in grado di differenziare i pazienti «fallers» dai «non fallers»

Il FES dimostra come il grado di confidenza risulta più basso nei pazienti con maggiore frequenze di cadute. FES non-fallers = 30.8 (25.1) e FES fallers = 40.4 (23.4)

(Cakit et al, 2007; $n = 31$ with diagnosis of idiopathic PD; mean age = 71.8 (6.4); duration of PD diagnosis = 5.58 (2.9) years; UPDRS motor subscale mean = 18.14 (9.32); MMSE > 20/30; tested in "on" state of medications. RCT with random assignment to treadmill walking or control group)

● ● ● | Berg Balance Scale

Obiettivo:

- 14 item creati per valutare l'equilibrio statico e il rischio di caduta nella popolazione adulta



Dominio ICF:

- Attività

<https://www.youtube.com/watch?v=HBKXu9fHnuo>

SEM

Anziani

Berg Balance Scale Initial Score	SEM
0 - 24	1.7
25 - 34	2.3
35 - 44	1.8
45 - 56	1.2

(Donoghue et al, 2009; $n = 118$ people over 65 years of age without a previous history of stroke)

Ictus

- Intero gruppo ($n = 48$) = 2.49
- Individui che deambulano con assistenza ($n = 16$) = 2.93
- Individui che deambulano con stand-by-assist ($n = 17$) = 2.15
- Individui che deambulano indipendentemente ($n = 15$) = 2.26

(Liston and Brouwer, 1996; $n = 22$ subjects with hemiparesis associated with unilateral stroke)

Trauma cranico

- Calcolato dalle statistiche degli articoli SEM = 1.65

(Newstead et al, 2005; $n = 5$; mean age = 24.4)

MDC

Anziani

Berg Balance Scale Initial Score	MDC
0 - 24	4.6
25 - 34	6.3
35 - 44	4.9
45 - 56	3.3

(Donoghue et al, 2009)

Parkinson

MDC = 5 punti

(Steffen and Seney, 2008; $n = 37$; mean age 71 years)

Ictus Acuto

- MDC per un intero gruppo($n = 48$): 6.9
- MDC per individui che non deambulano($n = 16$): 8.1
- MDC per individui che deambulano con assistenza ($n = 17$): 6.0
- MDC per individui che deambulano indipendentemente ($n = 15$): 6.3

(Stevensen, 2001; during inpatient rehabilitation)

CUT-OFF

Anziani

- Un punteggio di 56 indica l'equilibrio funzionale.
- Punteggio < 45 indica individui che hanno più probabilità di cadere.

(Berg et al, 1992; $n = 70$ acute stroke patients; mean age = 71.6 (10.1) years; $n = 113$ individuals from a home for the elderly; mean age = 83.5 (5.3) years; $n = 31$ elderly individuals who agreed to participate in a laboratory study; mean age = 83.0 (6.9) years;)

Lesione del midollo spinale

- Non è stata trovata nessuna relazione significativa tra le cadute totali e i punteggi BBS.
- Non è stato trovato nessun punteggio che discrimina le cadute.

(Wirz et al, 2010; $n = 42$ participants with SCI; mean age = 49.3 (11.5) years)

Ictus

- Punteggio = 45 su 56.

(Doggan et al, 2011; $n = 51$; mean age = 60.7)

Reliability

Test-retest Reliability

Anziani

Excellent test-retest reliability (ICC = 0.91)
(Berg et al, 1992; Elderly)

Parkinsonismo

Excellent test-retest reliability (ICC = 0.94)
(Steffen and Seney, 2008)

Ictus

Excellent test-retest reliability (ICC = 0.98)
(Liston and Brouwer, 1996)

Trauma Cranico

Excellent test-retest reliability (ICC = 0.986)
(Newstead et al, 2005)

Interrater/Intrarater Reliability

Anziani

Excellent intrarater reliability (ICC = 0.97)
(Berg et al, 1992; Community-Dwelling Elderly)

Lesione midollo spinale

Excellent interrater reliability (ICC = 0.95)
(Wirz et al, 2010; Chronic SCI)

Ictus

Excellent interrater reliability in individuals 14 days post (ICC = 0.95)
(Berg et al, 1995; n = 113 elderly residents and 70 stroke patients; mean age = 84.4(5.0) years; Acute Stroke)

Validity

Criterion Validity

Anziani

- Correlazione eccellente con il Dynamic Gait Index ($r = 0.67$)

(Shumway-Cook et al, 1997)

Ictus

- Correlazione eccellente con la sottoscala di equilibrio di Fugl-Meyer a 14, 30, 90 e 180 giorni dopo l'ictus ($r = 0.90$ a 0.92)
- Correlazione eccellente con la Postural Assessment Scale per pazienti colpiti da ictus (PASS) ($r = 0.92$ to 0.95)

(Mao et al, 2002; $n = 123$; mean age = 69.3 (11.2) years)

Construct Validity (Convergent/Discriminant)

Parkinson

BBS	50FW-S	LFIM	FIM	WISCI
3mo	Excellent 0.81*	Excellent 0.89	Excellent 0.76	Excellent 0.91*
6mo	Excellent 0.86	Excellent 0.86	Excellent 0.72	Excellent 0.89*
12mo	Excellent 0.78	Excellent 0.86	Excellent 0.77	Excellent 0.92*

$p < 0.001$

BBS – Berg Balance Scale

50FW-S – 50 foot walking speed

LFIM – locomotor Functional independence Measure

FIM- Functional Independence Measure

WISCI – Walking Index for Spinal Cord Injury

(Brusse et al. 2005; $n = 25$; mean age = 76 (7) years)

● ● ● | Timed Up and Go

Obiettivo:

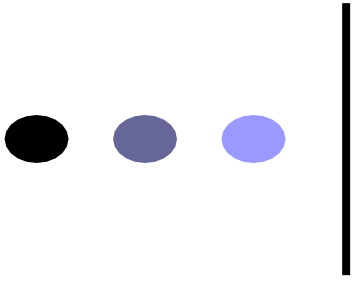
- Valutare mobilità, equilibrio, abilità del cammino e rischio di caduta nelle persone anziane.



https://www.youtube.com/watch?v=BA7Y_oLElGY

Dominio ICF:

- Attività



Le fonti suggeriscono che:

- un tempo di dieci secondi o meno indica una normale mobilità;
- tempi tra 11 e 20 secondi sono nei normali limiti per anziani con fragilità e pazienti disabili;
- tempi superiori ai 20 secondi indicano che la persona necessita di assistenza esterna e la necessità di ulteriori esami e interventi;
- un punteggio superiore ai 30 secondi suggerisce che la persona potrebbe essere soggetta a cadute.

SEM

<u>Parkinson</u>	<u>Lctus</u>	<u>Alzheimer</u>
SEM = 1.75 seconds (Dal BelloHaas et al, 2011; n = 24; mean age = 64.9 (8.0) years; mean time since diagnosis = 4.5 (4.3) years; H & Y Stage One = 13, Stage Two = 6, Stage Three = 5; mean MMSE scores = 27.4 (2.5) points, PD)	SEM = 1.14 seconds (Flansbjer et al, 2005; n = 50; mean age = 58 6.4) years; 646months poststroke;Swedish sample, Chronic Stroke)	•SEM = 2.48 seconds for all participants •SEM = 1.52 seconds for mild to moderate AD •SEM = 3.03 seconds for moderately severe to severe AD (Ries et al, 2009*)

(*) I risultati ottenuti rimandano a test sperimentali di TUG effettuati su un gruppo di 51 pazienti(criteri di inclusione: alta probabilità di AD, stabilità medica e deambulazione con o senza ausilio). I pazienti sono stati suddivisi in due sottogruppi sulla base di classificazione tabellare:
n=20 da lieve a moderato livello di demenza AD (4 o 5);
n=31 da moderatamente grave a grave livello di demenza AD (6 o 7).

MDC (minimal detectable change):

<u>Parkinson</u>	<u>Lctus</u>	<u>Alzheimer</u>
<ul style="list-style-type: none">•MDC = 4.85 seconds <p>(Dal BelloHaas et al, 2011, PD)</p> <ul style="list-style-type: none">•MDC = 3.5 seconds <p>(Huang et al, 2011; n = 72; mean age = 67.5 (11.6) years; mean baseline TUG = 11.8 seconds; DGI = 21.6 points; Chinese language sample, PD)</p> <ul style="list-style-type: none">•MDC = 11 seconds <p>(Steffen & Seney, 2008; n = 37, mean age= 71 (12); mean H&Y score = 2 (range=1–4); mean disease duration = 14 (6) years, PD)</p>	<ul style="list-style-type: none">•MDC=2.9 seconds <p>(calculated from Flansbjer et al, 2005) Smallest Real Difference (SRD) =23%</p>	<ul style="list-style-type: none">•MDC90 = 4.09 seconds <p>(Ries et al, 2009, Alzheimer's Disease)</p>

CUT-OFF:

Cut-Off Scores indicating risk of falls by population		
Population	Cut-Off score	Author
Community dwelling adults	> 13.5*	Shumway-Cook et al, 2000
Older stroke patients	> 14*	Andersson et al, 2006
Older adults already attending a falls clinic	> 15*	Whitney et al, 2005
Frail elderly	> 32.6*	Thomas et al, 2005
LE amputees	> 19*	Dite et al, 2007
Parkinson's Disease	> 11.5*	Nocera et al, 2013
	> 7.95*	Dibble et al, 2006
Hip Osteoarthritis	> 10*	Arnold et al, 2007
Vestibular Disorders	> 11.1*	Whitney et al, 2004
* Time in seconds		

Test-retest Reliability

<u>Parkinson</u>	<u>Ictus</u>	<u>Anziani</u>
<ul style="list-style-type: none">•Adeguato testretest reliability (ICC = 0.85) <p>(Steffen & Seney, 2008, Parkinson's Disease)</p> <ul style="list-style-type: none">•Eccellente testretest reliability (ICC = 0.80) <p>(Huang et al, 2011; n = 72 participants recruited from special clinics for movement disorders at a university hospital; sex = 44 males and 28 females, mean age = 67.5 (11.6) years; Taiwanese sample, Parkinson's Disease)</p>	<ul style="list-style-type: none">•Eccellente testretest reliability (ICC = 0.96) <p>(Flansbjer et al, 2005; 7 days between assessments, 16 to 18 months from stroke onset to initial assessment, Chronic Stroke)</p>	<ul style="list-style-type: none">•Adeguato testretest reliability per tutti i soggetti (ICC = 0.56) <p>(Rockwood et al, 2000; n = 2,305 elderly people, 874 males, 1431 females; mean age = 78.1 (69104) years; Canadian sample, Elderly Adults)</p>

	<u>Interrater</u>	<u>Intrarater</u>
Parkinson	Eccellente interrater reliability (ICC = 0.99) (Bennie et al 2003)	Eccellente intrarater reliability (ICC = 0.98) (Bennie et al 2003)
Anziani	Eccellente Interrater reliability (mean difference between raters = 0.04 seconds) (Siggeirsdottir et al, 2002; n = 31 elderly individuals living in a retirement home; median age = 82 (6592) years; Icelandic sample, Elderly Adults)	Non valutato

Predictive/Concurrent Validity:

<u>Parkinson</u>	<u>Lctus</u>	<u>Anziani</u>
<ul style="list-style-type: none">• Significativa correlazione tra TUG e BBS ($r=0.47$ e $p=0.04$) <p>(Bennie et al 2003).</p> <ul style="list-style-type: none">• Con un aumento di 16 sec associato ad un aumento di rischio di caduta ($p=0.043$) <p>(MAK e Pang, 2009).</p> <ul style="list-style-type: none">• Rischio di caduta: sensibilità 0.69; specificità 0.62; accuratezza: 0.63. <p>(Kerr et al, 2010).</p>	<ul style="list-style-type: none">• Eccellente correlazione tra TUG e CGS ($r = 0.86$)• Eccellente correlazione tra TUG e FGS ($r = 0.91$)• Eccellente correlazione tra TUG e SCas ($r = 0.86$)• Eccellente correlazione tra TUG e SCde ($r = 0.90$)• Eccellente correlazione tra TUG e 6MW ($r = 0.92$) <p>(Flansbjer et al, 2005, Chronic Stroke)</p>	<ul style="list-style-type: none">• Eccellente correlazione tra TUG e Berg Balance ($r = 0.81$)• Eccellente correlazione tra TUG e gait speed ($r = 0.61$)• Eccellente correlazione tra TUG e Barthel Index of ADL ($r = 0.78$) <p>(Podsiadlo & Richardson 1991; n = 60 patients referred to a geriatric day hospital; mean age = 79.5 years)</p> <ul style="list-style-type: none">• Il TUG è in grado di prevedere i rischi di caduta: Sensitivity 56%, Specificity 60% <p>(Bhatt T et al, 2011)</p>

Convergent/Discriminant Validity

<u>Parkinson</u> <u>(Convergent Validity)</u>	<u>Anziani</u> <u>(Convergent Validity)</u>	<u>Geriatrici</u> <u>(Discriminant Validity)</u>
<ul style="list-style-type: none"> •BBS $r=0.78$ (Eccellente) •FGS $r=0.69$ (Eccellente) •CGS $r=0.67$ (Eccellente) (Significativo per $p < 0.001$) <p>(Brusse et al, 2005, Parkinson's Disease)</p>	<p>Adeguate: TUG e Tinetti Balance ($r = 0.55$)</p> <p>Adeguate: TUG e Tinetti Gait ($r = 0.53$)</p> <p>Adeguate: TUG e walking speed ($r = 0.66$)</p> <p>Adeguate: TUG e ADL scale ($r = 0.45$)</p> <p>(Lin et al, 2004; $n = 1200$; mean = 73.4 years; Taiwanese sample, Community Dwelling Older Adults)</p>	<ul style="list-style-type: none"> •Adeguate correlazione tra TUG e Functional reach ($r = 0.36^*$) •Eccellente correlazione tra TUG e 2MWT ($r = 0.68^*$) <p>* TUG Correlations are expected to be negative, lower scores equal better outcomes</p> <p>(Brooks et al, 2006)</p>



Confronto tra FES, BBS e TUG

Anziani

	FES	BBS	TUG
SEM	-----	2.3	3,03 s
MDC	-----	6.3	-----
Reliability	Adeguate	Eccellente	Adeguate
Criterion Validity	Eccellente	Eccellente con il DGI	Eccellente BBS

Ictus

	FES	BBS	TUG
SEM	-----	2.93	1.14 s
MDC	-----	8.1	2.9 s
Reliability	Eccellente	Eccellente	Eccellente
Criterion Validity	-----	Eccellente con PASS e FM	Eccellente con SMW