

**100** 1925—2025  
UN INIZIO  
STRAORDINARIO



**Ambasciatrice dell'italiano  
nel mondo**





CENTRO VALUTAZIONE  
CERTIFICAZIONI LINGUISTICHE  
*Università per Stranieri Perugia*



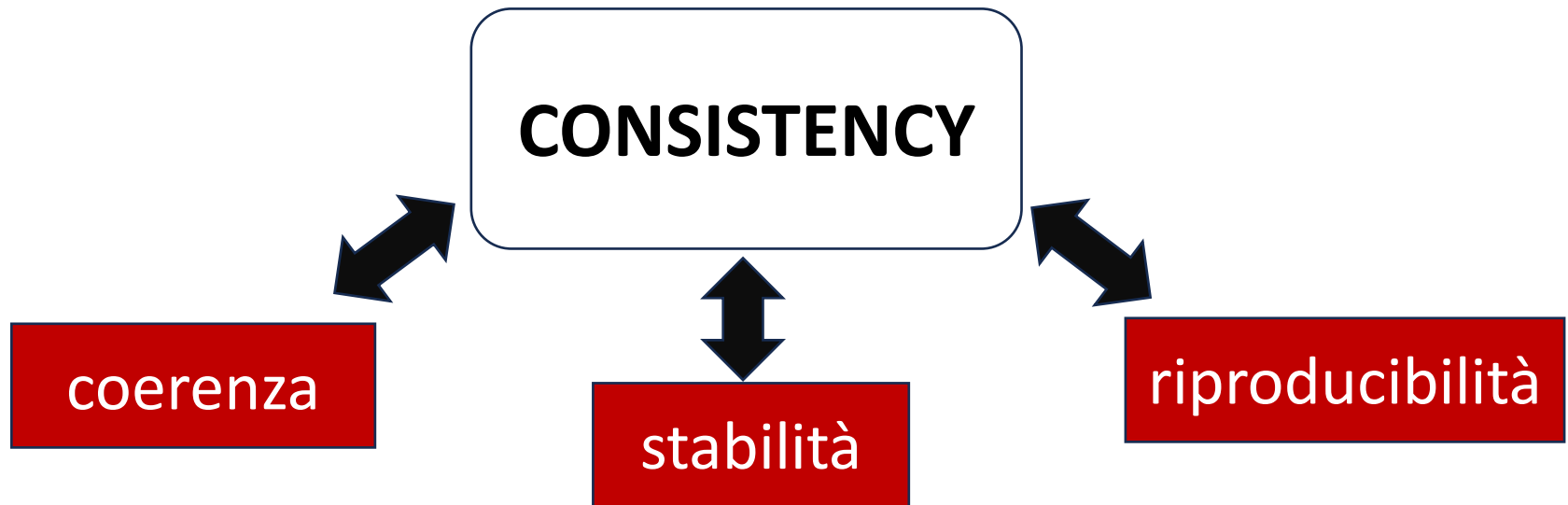
# **Principi fondanti di una valutazione e di un testing di qualità (validità, affidabilità, praticabilità e benefici per gli stakeholders)**

Formatore: Claudio Battaglia

[claudio.battaglia@unistrapg.it](mailto:claudio.battaglia@unistrapg.it)

# Che cosa è un test?

- I TEST sono forme di **misurazione** numerica della conoscenza, delle abilità e delle competenze (**KSA**) linguistiche, esibite da un individuo. La misurazione può essere espressa in modalità assoluta o per gradazioni. (Greenwood, 1919).
- «It is assumed that tests, like scientific instruments, provide the means by which we can observe and gauge **consistencies** in human ability». (G. Fulcher, *Practical Language Test*)

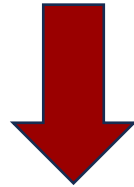


# **SUPERPOTERI**

- **Immaginate di avere un superpotere, questo superpotere vi consente di decidere se rimuovere tutte le forme di testing e esami che sono state create dalla società umana nel corso dei secoli. Una volta che avrete utilizzato il vostro potere ogni forma di test ed esame sparirà e non potrà più essere recuperata. Prima di decidere se utilizzare il vostro superpotere oppure no, pensate a varie forme di test o prove utilizzate nella società (non solo test linguistici), provate ad individuarne e a scriverne almeno 3. Poi per ciascuna delle vostre opzioni provate a dire anche come eventualmente potreste ottenere gli stessi risultati (e misurazioni) garantiti dal test o dall'esame, senza però utilizzare il test.**
- **Alla fine decidete se valga la pena usare il vostro potere e cancellare definitivamente ogni forma di test dalla faccia della terra.**

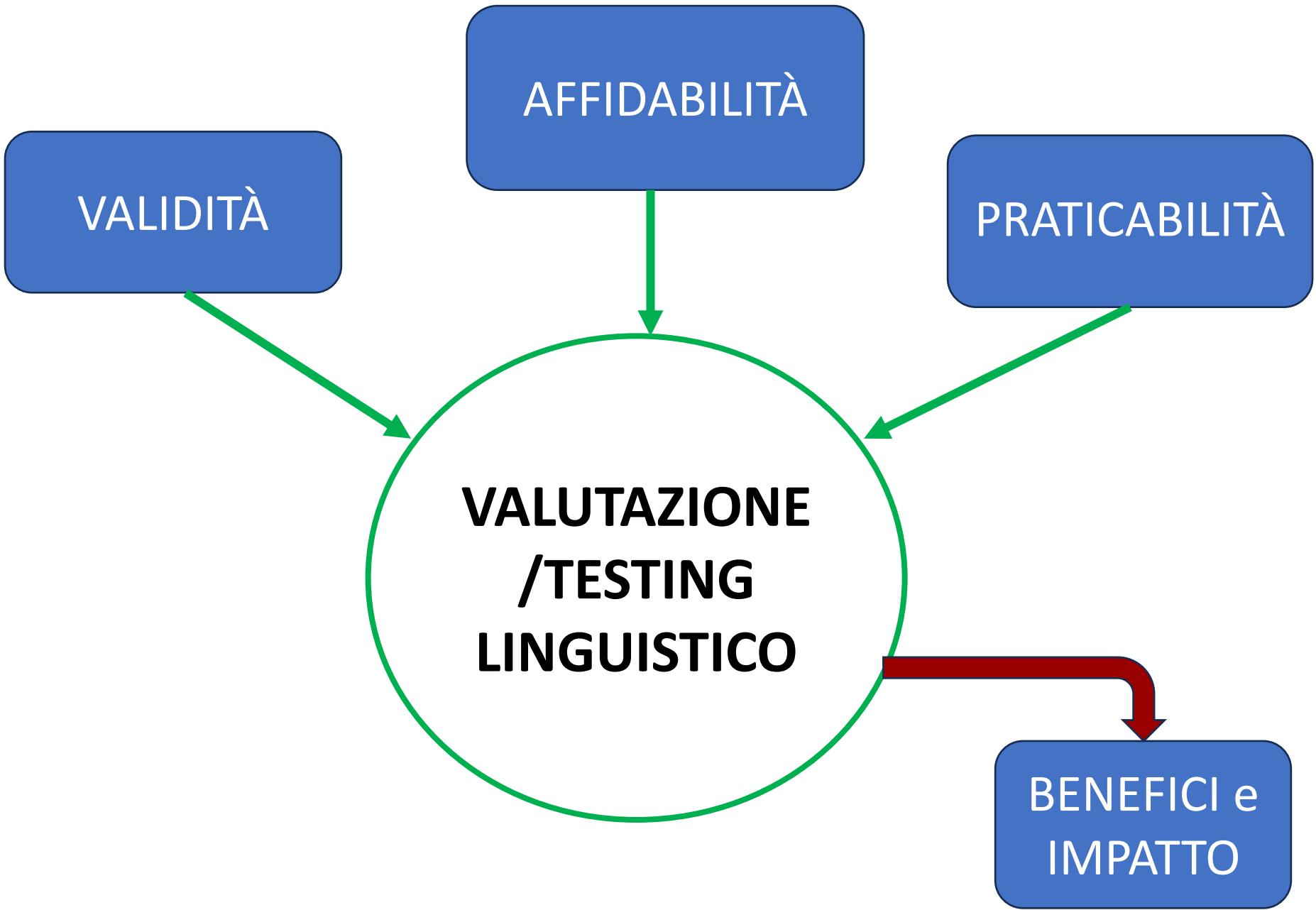
# La lingua può essere valutata con precisione?

- È fondamentale ricordare che la misurazione (e conseguentemente la valutazione) di KSA linguistiche implica un certo grado di approssimazione (o incertezza) e anche di errore nella misurazione stessa.



La valutazione (o misurazione) in campo linguistico consiste nel connettere indissolubilmente concetti astratti, come le KSA, ad indicatori empirici e quantificabili (i task presenti in un test ed eseguiti dai candidati).

- È possibile raccogliere ed utilizzare diversi tipi di evidenze collegate alla validità di un processo di valutazione/misurazione, per dimostrare che i risultati ottenuti in un test sono, appunto, validi.



VALIDITÀ

AFFIDABILITÀ

PRATICABILITÀ

**VALUTAZIONE  
/TESTING  
LINGUISTICO**

BENEFICI e  
IMPATTO

Il termine **VALIDITÀ** si riferisce alla qualità di un test nel raggiungere con precisione l'oggetto del processo di valutazione senza confonderlo con fattori o elementi non pertinenti. Più alto è il grado con cui le evidenze raccolte e la teoria di acquisizione della lingua di riferimento supportano l'interpretazione dei risultati del test in relazione alle finalità e agli scopi per cui il test è stato realizzato, più alta è la validità del test.

**La validità non è propriamente una qualità intrinseca del test ma esprime il grado con cui le evidenze e l'assunto teorico supportano l'interpretazione dei punteggi raggiunti dai candidati, interpretazione esplicitamente dichiarata nelle finalità del test.**

# i diversi ambiti in cui rilevare evidenze a sostegno della VALIDITÀ

## VALIDITÀ basata sul CONTENUTO

È prodotta dalla corrispondenza tra i contenuti e le caratteristiche del test e dei compiti da esso previsti e:

- Il tipo di lingua/aspetti ed elementi linguistici da valutare
- Le conoscenze, abilità e competenze da valutare

## VALIDITÀ basata sul CRITERIO

È prodotta dalla corrispondenza tra la performance in un test e i descrittori/indicatori presenti in un quadro di riferimento (QCER o altri). La corrispondenza può anche stabilirsi tra la performance in un test e la valutazione di un team di esperti esterno al contesto di apprendimento o valutazione in cui il test è stato somministrato

## VALIDITÀ basata sul COSTRUTTO

È l'aspetto della validità che mette in relazione i due precedenti e si riferisce all'impianto teorico (teoria linguistico-acquisizionale e cognitiva di riferimento) su cui si basa il sistema di valutazione predisposto

**VALIDITÀ di  
ASPETTO (*face  
validity*)**



È relativa alla misura in cui un test «appare», per il modo in cui è concepito e strutturato, come una misura valida di ciò che si propone di misurare. È un tipo di validità accordata dagli stakeholders (apprendenti, impiegati, corpo docente, rappresentanti istituzionali, che non sono necessariamente esperti di lingua e valutazione). È strettamente correlata ai task presenti nel test e al tipo di costrutti ad essi collegati.

**La tendenza più diffusa oggi, in campo scientifico, è quella di considerare la validità come un concetto unitario, espressione del cosiddetto processo di validazione di un test, cioè della raccolta e dell'organizzazione di evidenze per supportare l'interpretazione dichiarata o l'uso previsto dei risultati ottenuti in un determinato test (o prova di certificazione linguistica).**

È fondamentale, per il funzionamento ottimale (validità) di un sistema di valutazione o test, che tutti coloro che sono coinvolti, a vario titolo, nell'interpretazione dei risultati del test stesso, condividano una conoscenza generale del **costrutto** o dei **costrutti** selezionati nel progettare e nel creare il test.

La validità di un sistema di valutazione può essere minata dalla presenza di:

- 1. costrutti irrilevanti**
- 2. costrutti sottorappresentati**

- **irrilevanza di costrutto**

- Un test per la misurazione dell'abilità di produzione scritta che prevede una prova di ascolto con task a risposta a scelta multipla.

- **costrutto sottorappresentato**

- Un test che si propone di misurare il livello di competenza linguistico-comunicativa globale dei candidati ma che non preveda alcuna prova di ascolto.

# TASK 1

- Siamo in Italia. In un CPIA (Centro Provinciale per l'Istruzione agli Adulti) l'esame di fine anno per i corsi di italiano L2 di livello A2 del QCER (riconosciuti dalla normativa vigente per il rilascio del permesso di soggiorno di lungo periodo) prevede una prova scritta con task sulla competenza linguistica, sulla produzione scritta, sulla comprensione della lettura e sulla comprensione dell'ascolto. La competenza di produzione/interazione orale è valutata sulla base dell'osservazione condotta dal docente durante lo svolgimento delle lezioni. Quindi in sede d'esame finale non sono previste prove ufficiali per la valutazione della competenza di produzione/interazione orale.
- Che problemi si presentano nel garantire la validità collegata ad un tipo di test/esame come questo?

# AFFIDABILITÀ

- Il termine **AFFIDABILITÀ** si riferisce alla proprietà di un test di presentare gli stessi dati ed esiti (invariabilità/costanza), pur variando il contesto di erogazione e i somministratori o le commissioni d'esame. Che cosa valutiamo? Come è valutata la performance del candidato?
- La principale minaccia all'affidabilità di un test è rappresentata dagli:

**ERRORI DI MISURAZIONE**

Gli errori di misurazione possono essere collegati a:

## **1. Tempo**

## **2. Candidati/studenti**

- caratteristiche fisiologiche
- caratteristiche psicologiche
- caratteristiche esperienziali

## **3. Esaminatori/valutatori**

- bias soggettivi nel processo di valutazione (errori *casuali* o *sistematici*)

Esiste un modello di riferimento per garantire l'**affidabilità** di un test (Hughes 2003, Brown 2005, Douglas 2010):

- 1. Produrre test chiari e non ambigui nelle consegne e negli obiettivi**
- 2. Impiegare diversi tipi di testlet e task**
- 3. Delimitare chiaramente ciò che viene valutato e perché viene valutato**
- 4. Standardizzare le condizioni di somministrazione e valutazione**
- 5. Controllare come viene corretto il test e come vengono assegnati i punteggi, predisponendo anche un vademecum per il *marking***
- 6. Avvalersi di più correttori, tutti adeguatamente formati**
- 7. Valutare i candidati in un ampio ventaglio di abilità e competenze**

## TASK 2

**Nel dipartimento di italianistica di un'università straniera un docente incaricato ha predisposto un test per valutare le abilità di ascolto di livello A2 che conta il 25% del punteggio totale nell'esame finale. Il test consiste *nell'ascolto di una conversazione in hotel tra un'ospite appena arrivata e il receptionist, durante l'ascolto i candidati devono riempire una scheda con i dati dell'ospite e le necessità da lei espresse.***

**Dato che questa è l'unica forma di test di ascolto impiegata per il livello A2 il docente non ha predisposto test-spec, linee guida o materiali extra per i suoi colleghi. Gli unici materiali prodotti e condivisi sono il file audio e il foglio per svolgere il task. Ogni insegnante di italiano del dipartimento somministra il test in classe, valuta individualmente le risposte (su una scala che prevede il punteggio massimo di 12, già adottata per un altro formato d'esame) e consegna i risultati alla segreteria del dipartimento.**

## **TASK 2**

**Lavorando in piccoli gruppi provate a rispondere alle domande.**

- 1. Che tipo di problemi possono sorgere che vadano a compromettere l’AFFIDABILITÀ del test?**
- 2. Che tipo di cambiamenti proporreste per aumentare l’AFFIDABILITÀ del test?**

# PRATICABILITÀ

«Practicality is the difference between the resources that will be required in the development and use of an assessment and the resources that will be available for these activities».

*(Bachman & Palmer, 2010)*

Il termine **PRATICABILITÀ** si riferisce, nello specifico, alla proprietà legata alle caratteristiche procedurali del sistema di valutazione, sia in termini di risorse (umane, economiche e strumentali) e spazi a disposizione dei candidati e degli enti esaminatori, sia in termini di complessità dei criteri e delle categorie impiegati nel processo valutativo, con cui gli esaminatori hanno a che fare nel valutare una determinata performance.

- La praticabilità fa riferimento al fabbisogno di risorse (materiali e umane) necessarie all'implementazione di un esame o di un sistema certificadorio, cioè alla sua **fattibilità** e alla sua **sostenibilità**. È strettamente legata a:
  - Numeri
  - Livello di formazione e specializzazione del personale
  - Spazi e loro gestione
  - Materiali
  - Tempistiche
  - Strumenti e supporti a disposizione
  - Budget

## BENEFICI E RICADUTE PER GLI STAKEHOLDERS

Un test, un esame e, ancora di più, una certificazione linguistica possono rappresentare un meccanismo più equo, giusto, sostenibile e inclusivo nel decretare l'accesso ad opportunità rilevanti in settori come l'immigrazione, l'istruzione, il lavoro. Ma non sono esenti da falle. È possibile che il contenuto, il formato e le modalità di somministrazione di un esame lavorino, in modo sistematico, a favore di un gruppo di candidati e a sfavore di un altro.

Un sistema certificatorio realmente valido e utile è quello che produce benefici e non complicazioni per tutti gli stakeholders.

# IMPATTO

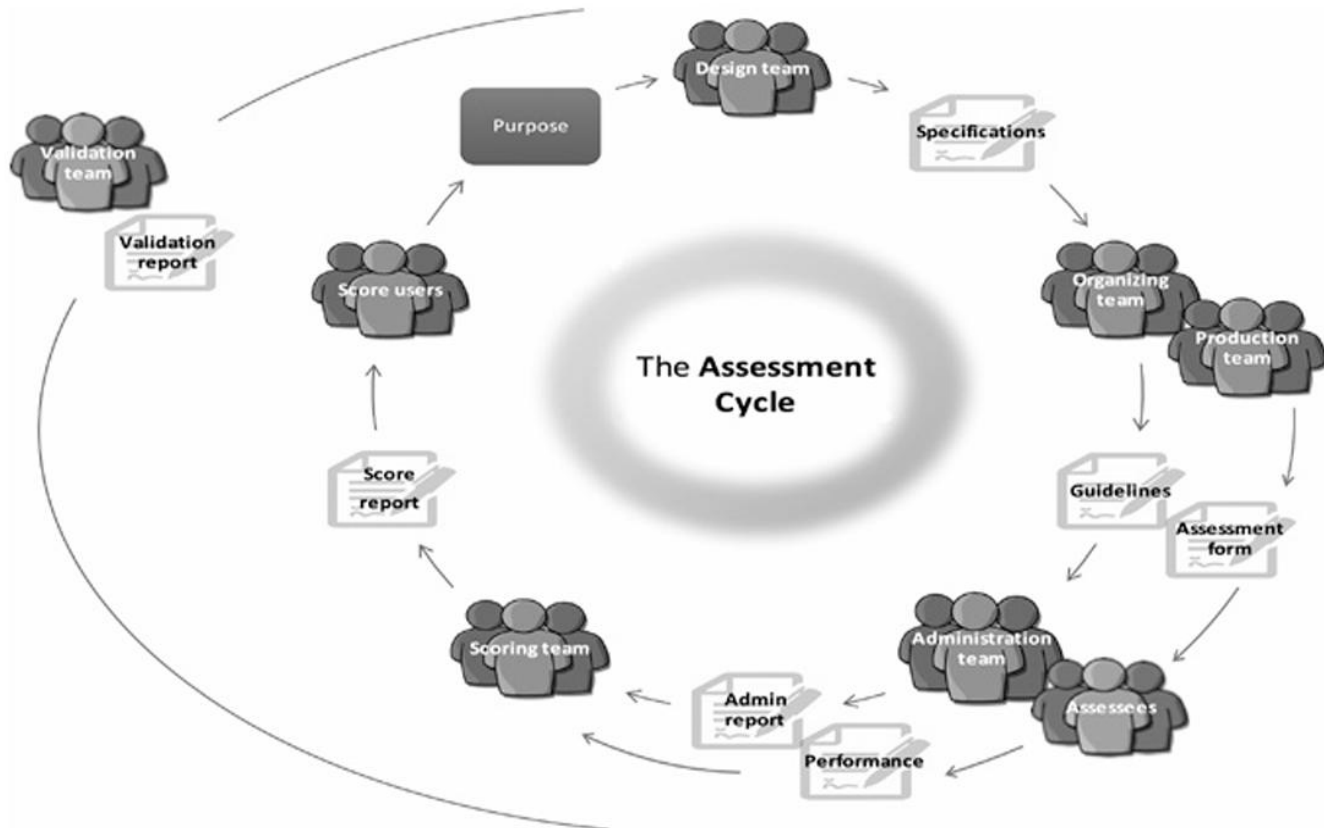
- **AMBITO SOCIALE:** conseguimento della cittadinanza e maggiore integrazione socio-economica.
- **AMBITO FORMATIVO E PROFESSIONALE:** accesso a università, percorsi accademici e di alta formazione, opportunità professionali.
- **FEEDBACK/WASHBACK:** processi legati alle dinamiche di insegnamento/apprendimento linguistico in classe.
- ...

# EQUITÀ

- È fondamentale che gli enti certificatori minimizzino effetti negativi presenti nel test e causati da bias o da bassi livelli di inclusione: nella scelta degli argomenti, nell'attenzione ai bisogni educativi speciali (BES) e ai disturbi specifici dell'apprendimento (DSA), nelle procedure di somministrazione e valutazione del test.
- La garanzia dei criteri di eticità e di equità del testing è oggi ancora più stringente e attuale, considerando le sfide legate all'avvento dell'IA e al suo rapido e continuo sviluppo.

# In che modo gli esami CELI garantiscono i criteri di **VALIDITÀ, AFFIDABILITÀ e PRATICABILITÀ**?

The practice of language assessment



## Riferimenti bibliografici e sitografici

- Fulcher, G., Harding, L., (eds.) (2022). «The Routledge Handbook of Language Testing». London UK: Routledge.
- Fulcher, G., (2010). «Practical Language Testing». London UK: Routledge.
- Grego Bolli, G., Spiti, M. G., (2004). «Misurare e valutare nella certificazione àCELL: linee guida alla certificazione dell'Università per Stranieri di Perugia/Nuova ediz. riv e ampliata». Perugia: Guerra.
- Green, A. (2021, 2° ed.). «Exploring Language Assessment and Testing (Language in Action)». London UK: Routledge.
- Piccardo, E., Berchoud, M., Cignatta, T., Mentz, O., Pamula, M. (2011). «Pathways through assessing, learning and teaching in the CEFR». Pp. 42 - 53. Consiglio d'Europa [Microsoft Word - 2011\\_08\\_28\\_ECEP.doc](#)
- Barsi, M., Lugarini, E., & Cardinaletti, A. (a cura di). (2020). «Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione. Volume complementare». Consiglio d'Europa.

## Riferimenti bibliografici e sitografici

- «Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)». [Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment \(CEFR\). A Manual - Common European Framework of Reference for Languages \(CEFR\)](#)
- [Manual for Language Test Development and Examining](#)



CENTRO VALUTAZIONE  
CERTIFICAZIONI LINGUISTICHE

*Università per Stranieri Perugia*



**Grazie!**