

GLI APPROFONDIMENTI DEL SERVIZIO STATISTICO

La newsletter ufficiale del Servizio Statistico INVALSI



Gentilissimo/a,

la newsletter *“Gli approfondimenti del Servizio Statistico”* del mese di ottobre è dedicata all’equità nella valutazione degli apprendimenti. Prima ancora di leggere i punteggi e interpretare le differenze tra studenti e studentesse, classi e territori, occorre interrogarsi sulla qualità degli strumenti con cui quei risultati vengono rilevati.

Equità nella valutazione

Misurare bene per comprendere le disuguaglianze educative

Che cosa significa costruire una prova che misuri in modo imparziale?

Come possiamo individuare quesiti che funzionano diversamente per gruppi di studenti e studentesse?

E in che modo una misurazione accurata può aiutarci a comprendere le disuguaglianze educative?

In questo numero seguiremo il percorso che collega la costruzione delle prove, l'interpretazione dei risultati e l'utilizzo delle informazioni per orientare il miglioramento.

PRIMA DEI RISULTATI: LA QUALITÀ DELLA MISURAZIONE

Quando si affronta il tema della valutazione degli apprendimenti, l'attenzione si concentra quasi esclusivamente sui risultati conseguiti. Ci si interroga, infatti, su punteggi di studenti e studentesse, divari territoriali, differenze tra classi e scuole e cambiamenti nel tempo. Molto più raramente ci si sofferma però su una domanda che precede tutte le altre: ci si può davvero fidare di ciò che si sta misurando?

La domanda non è soltanto tecnica. Se uno strumento di valutazione introduce, anche involontariamente, elementi che favoriscono alcuni studenti e studentesse o ne penalizzano altri e altre, il risultato osservato può restituire un'immagine distorta di ciò che si intende misurare. Una prova può essere costruita con grande attenzione, essere coerente con il quadro di riferimento e funzionare bene dal punto di vista statistico: questo però non significa che sia necessariamente imparziale nei confronti dei diversi gruppi di studenti e studentesse.

È qui che entra in gioco un'altra dimensione della qualità della misurazione. Una valutazione, oltre a dover essere affidabile e valida, deve consentire a studenti e studentesse di mettere in gioco gli apprendimenti che si intende osservare, senza che fattori estranei all'oggetto della prova introducano ostacoli non pertinenti.

Tutto questo porta a una questione ancora più importante: che cosa significa, concretamente, misurare in modo imparziale? Significa poter interpretare le differenze nei risultati con una ragionevole fiducia, sapendo che non dipendono da caratteristiche della prova estranee alla manifestazione dell'apprendimento che vogliamo misurare. Ciò non implica, naturalmente, che tutti gli studenti e tutte le studentesse debbano ottenere lo stesso punteggio.

È una distinzione tutt'altro che secondaria. Se, ad esempio, una prova di matematica, ancor prima di richiedere un determinato calcolo, necessita di comprendere un testo linguisticamente molto complesso, una parte della difficoltà potrebbe dipendere dalla comprensione linguistica e non dalla competenza matematica. In questo caso il punteggio ottenuto non sarebbe necessariamente una fotografia fedele di ciò che si voleva osservare.

Per questa ragione, l'attenzione all'imparzialità della misurazione non può essere considerata un elemento aggiuntivo della valutazione, da affrontare una volta terminata la costruzione della prova. È una condizione che riguarda l'intero processo di misurazione e, in ultima analisi, la possibilità stessa di interpretare correttamente i risultati.

EQUITÀ NON SIGNIFICA UGUAGLIANZA

Uno dei primi equivoci da evitare è quello di considerare uguaglianza ed equità come sinonimi. L'uguaglianza richiama l'idea di offrire a tutti e tutte le stesse condizioni: la stessa prova, le stesse istruzioni, gli stessi tempi, le stesse modalità di svolgimento. In una valutazione standardizzata questo principio è essenziale, perché senza condizioni comparabili sarebbe difficile confrontare i risultati.

Ma è sufficiente garantire a tutti e tutte le stesse condizioni di somministrazione perché la misurazione sia imparziale? Non necessariamente. Gli studenti e le studentesse si avvicinano alla prova partendo da situazioni, condizioni e percorsi differenti e molto eterogenei. Hanno avuto esperienze scolastiche diverse, vivono in contesti sociali e culturali differenti e possono avere una familiarità molto diversa con determinate forme linguistiche, situazioni o modalità di comunicazione. Questo non significa che tali differenze debbano essere eliminate dalla valutazione. Significa però che occorre distinguere ciò che appartiene all'apprendimento che vogliamo misurare da ciò che, invece, rischia di interferire con la sua rilevazione.

Il punto, dunque, non è fare in modo che tutti gli studenti e le studentesse conseguano lo stesso risultato; si tratta invece di fare in modo che una differenza di risultato possa essere interpretata come una differenza nell'apprendimento misurato e non, almeno in misura significativa, come l'effetto di un ostacolo estraneo a ciò che la prova intende rilevare.

Torniamo all'esempio del quesito di matematica. Se la complessità linguistica è parte integrante dell'apprendimento che si vuole rilevare, non sarebbe corretto considerarla un elemento estraneo. Se invece la lingua rappresenta soltanto un veicolo per accedere a un problema matematico, una formulazione inutilmente complessa potrebbe introdurre una difficoltà che non appartiene all'oggetto della misurazione.

Si ha dunque bisogno di condizioni standard che rendano possibili i confronti. Allo stesso tempo, si deve evitare che condizioni formalmente identiche si traducano in ostacoli diversi per studenti e studentesse, quando questi ostacoli sono estranei a quello specifico apprendimento oggetto della prova.

L'attenzione alla qualità della misurazione permette di predisporre domande in grado di verificare in modo adeguato e corretto i livelli di apprendimento raggiunto senza introdurre ostacoli impropri rispetto a ciò che si intende osservare. Per interpretare un risultato è quindi necessario comprendere quali richieste della prova siano pertinenti e quali possano introdurre una difficoltà aggiuntiva non prevista.

Equità non significa uguaglianza



COME SI COSTRUISCE UNA PROVA IMPARZIALE

L'imparzialità di una prova si costruisce fin dalla progettazione e va verificata lungo tutto il processo, prima ancora di interpretare i risultati. Dietro una prova non ci sono semplicemente una serie di domande a cui rispondere: ci sono un quadro di riferimento, una definizione degli obiettivi di apprendimento da rilevare, scelte linguistiche e contenutistiche, revisioni, sperimentazioni e analisi statistiche. Ogni passaggio può contribuire alla qualità della misurazione, ma anche introdurre, se non adeguatamente controllato, qualche elemento di distorsione.

Già nella fase di progettazione gli autori e le autrici devono chiedersi se il quesito rilevi effettivamente ciò che si intende misurare. Il linguaggio deve essere comprensibile ma non necessariamente semplificato: la questione è capire quale ruolo abbia la componente linguistica rispetto alla competenza oggetto della prova. Allo stesso modo, un riferimento culturale non è automaticamente problematico: lo diventa quando introduce una conoscenza che non dovrebbe essere necessaria per rispondere correttamente alla domanda.

Dalla costruzione dei quesiti all'imparzialità della prova

L'imparzialità si costruisce prima della somministrazione



L'imparzialità è parte della progettazione della prova e va verificata lungo tutto il processo.

Proprio per questa ragione, la revisione da parte di esperti ed esperte con competenze e professionalità diverse risulta cruciale: uno sguardo disciplinare può non essere sufficiente a individuare una formulazione ambigua, un riferimento culturale troppo specifico o una caratteristica del quesito che potrebbe avere conseguenze inattese per alcuni gruppi di studenti.

La sperimentazione è un passaggio essenziale: la prova si confronta con le risposte di studenti e studentesse reali. Le analisi statistiche dei pre-test permettono di osservare la difficoltà dei quesiti, la loro capacità di distinguere livelli diversi di competenza e gli eventuali comportamenti inattesi, anche tra gruppi differenti.

Nessuna procedura, presa singolarmente, può escludere la presenza di possibili distorsioni. La combinazione di osservazioni qualitative e quantitative permette di costruire un quadro più solido. Le evidenze raccolte durante lo sviluppo e l'utilizzo dello strumento possono rendere necessarie nuove verifiche e revisioni.

IL FUNZIONAMENTO DEI QUESITI E IL RUOLO DEL DIF

Un quesito può essere corretto dal punto di vista disciplinare e presentare comunque un funzionamento differente tra gruppi di studenti e studentesse. Può risultare più difficile per un gruppo rispetto a un altro anche quando i due gruppi presentano livelli comparabili nell'oggetto misurato. Occorre allora approfondire l'origine della differenza.

Uno degli strumenti utilizzati è l'analisi del *Differential Item Functioning* (DIF), ossia il funzionamento differenziale dell'item. L'analisi individua quesiti che funzionano diversamente per gruppi di studenti e studentesse, a parità di competenza stimata.

L'idea di fondo è intuitiva: se due studenti o studentesse hanno lo stesso livello nell'apprendimento che la prova intende rilevare, ci si aspetta probabilità simili di rispondere correttamente a uno stesso quesito. Una differenza tra gruppi richiede di certo attenzione ma non dimostra automaticamente la presenza di una distorsione.

La differenza può dipendere da un ostacolo non pertinente introdotto dal quesito oppure da una caratteristica che appartiene all'apprendimento stesso. Il risultato statistico va quindi interpretato alla luce del contenuto dell'item, del costrutto misurato e delle modalità con cui viene affrontata la prova.

Anche la "parità di apprendimenti" richiede cautela: si riferisce all'obiettivo di apprendimento definito e stimato attraverso il modello di misurazione. La verifica di possibili distorsioni riguarda pertanto sia il funzionamento del quesito sia la qualità della misura utilizzata per il confronto.

LA COMPARABILITÀ DEI RISULTATI

Quando una prova viene utilizzata su larga scala, non interessa soltanto contare le risposte corrette. Si vogliono interpretare i risultati, confrontarli tra studenti e studentesse e, in molti casi, osservare come cambiano gli apprendimenti nel tempo.

Per farlo occorrono scale di misura che consentano di collocare studenti e studentesse e quesiti in un sistema interpretativo coerente. I modelli statistici della misurazione educativa contribuiscono a costruire queste scale. La confrontabilità dei risultati richiede però verifiche sulle proprietà della misura e sul collegamento tra prove utilizzate in momenti o contesti diversi. Anche le condizioni di risposta devono essere considerate. La diffusione delle prove computer-based offre nuove possibilità di rilevazione e standardizza molte procedure, ma studenti e studentesse possono avere una familiarità diversa con gli strumenti digitali e con le modalità di interazione richieste.

È necessario chiedersi se queste differenze possano interferire con la rilevazione di quello specifico apprendimento. I nuovi strumenti digitali, che di certo possono contribuire alla qualità della valutazione, non rendono superflua la verifica di possibili fonti di distorsione. La comparabilità richiede quindi condizioni sufficientemente standardizzate per sostenere confronti affidabili, insieme all'attenzione necessaria a distinguere le differenze nelle modalità di risposta da quelle nella competenza misurata.

Scale di misura e comparabilità



La comparabilità è un requisito per un monitoraggio significativo dell'equità.

RENDERE VISIBILI LE DISUGUAGLIANZE EDUCATIVE

Che cosa accade quando una misurazione accurata restituisce un'immagine profondamente disuguale del sistema educativo?

Una buona misurazione può rendere più visibili le disuguaglianze, senza che questo segnali necessariamente una distorsione dello strumento. Al contrario, quanto più la misurazione è accurata e imparziale, tanto più le differenze osservate possono essere interpretate come informazioni sugli apprendimenti.

I risultati delle valutazioni su larga scala mostrano che le opportunità di apprendimento non sono distribuite uniformemente. Le condizioni socio-economiche e culturali delle famiglie sono associate agli apprendimenti e le differenze possono assumere forme diverse a seconda dei territori, delle scuole e dei gruppi di studenti.

Descrivere queste associazioni non implica di certo attribuire responsabilità agli studenti e alle studentesse o alle loro famiglie perché il focus sta nell'opportunità di comprendere quanto le condizioni nelle quali si cresce e si apprende siano collegate alle opportunità educative.

Il dato assume qui un significato che va oltre il confronto dei punteggi. Occorre comprendere dove si collocano le differenze, quanto sono ampie, come cambiano nei diversi contesti e quali meccanismi possono contribuire a produrle.

La variabilità, per esempio, può essere osservata soprattutto tra studenti e studentesse all'interno della stessa scuola oppure tra scuole e territori differenti. Sono fenomeni diversi, che richiedono interpretazioni diverse e che pongono domande diverse a chi deve intervenire.

Anche per questo una lettura dei risultati basata esclusivamente sulle classifiche rischia di essere riduttiva e di portare a effetti indesiderati. Un ordinamento può mostrare una differenza ma difficilmente può spiegare perché quella differenza esista.

L'imparzialità della misurazione e l'equità del sistema educativo sono dunque due questioni distinte ma strettamente collegate. La prima riguarda la qualità con cui vengono rilevati determinati apprendimenti; la seconda riguarda la distribuzione delle opportunità di apprendimento. Una misurazione accurata e imparziale è importante proprio perché permette di interrogarsi sull'equità del sistema educativo sulla base di informazioni più attendibili.

DALLA MISURAZIONE ALL'AZIONE

A questo punto, però, emerge una nuova domanda: cosa fare con ciò che la valutazione ci mostra? Produrre dati accurati è importante ma non sufficiente. Una misurazione può essere tecnicamente impeccabile ma avere un impatto nullo o limitato se le informazioni che produce non vengono comprese, discusse e utilizzate.

La qualità della misurazione e l'equità educativa riguardano dunque momenti diversi ma collegati: la prima entra in gioco nella costruzione della prova e nell'interpretazione dei risultati; la seconda chiama in causa ciò che quei risultati ci dicono sulla distribuzione delle opportunità di apprendimento. Se emergono differenze persistenti tra gruppi, territori o scuole, è necessario comprenderne il rapporto con le opportunità educative e interrogarsi sugli interventi che potrebbero contribuire a ridurle.

Questo richiede di tenere insieme livelli diversi: la qualità della misurazione, l'interpretazione dei dati, la conoscenza dei contesti e la capacità di trasformare le evidenze in decisioni.

È forse questo il punto in cui la valutazione mostra più chiaramente il suo significato. Misurare in modo accurato e imparziale non significa aspettarsi risultati uguali, ma creare le condizioni perché le differenze che osserviamo siano interpretabili e, quindi, perché possiamo interrogarci sulle disuguaglianze che quei dati rendono visibili.

Due prospettive sull'equità

Due domande, un unico sistema di valutazione



MISURAZIONE

- ✓ I quesiti sono imparziali?
- ✓ I punteggi sono comparabili?
- ✓ La somministrazione è standardizzata?
- ✓ L'errore di misurazione è controllato?



SISTEMA EDUCATIVO

- ✓ Le opportunità educative sono eque?
- ✓ I divari negli apprendimenti persistono?
- ✓ Dove si concentrano le disuguaglianze?
- ✓ Quali contesti richiedono interventi?

Valutazione



Evidenze



Azione

Nelle scuole, la lettura dei risultati può sostenere l'autovalutazione e la progettazione del miglioramento. A livello di sistema, può contribuire a individuare i contesti nei quali approfondire i bisogni educativi e orientare gli interventi. In entrambi i casi, l'interpretazione richiede di collegare i dati alla conoscenza dei contesti.

La valutazione, da sola, non può risolvere le disuguaglianze educative; può, tuttavia, contribuire a conoscerle meglio. Disporre di misure attendibili è perciò anche una responsabilità: richiede attenzione al modo in cui interpretiamo le differenze tra studenti e studentesse e utilizziamo le informazioni prodotte.

Rimane quindi centrale ciò che si sceglie di osservare attraverso la misurazione e come si decide di operare dopo avere interpretato in modo accurato e responsabile ciò che è emerso.

**Per seguire le attività e gli approfondimenti del
Servizio Statistico INVALSI:
<https://serviziostatistico.invalsi.it/>**

**A questo link è possibile recuperare
i precedenti numeri della newsletter**