

Ci soffermeremo quindi sui principali requisiti della validità (*validity*) e dell'affidabilità (*reliability*), sottolineando quanto le questioni non siano soltanto o principalmente di ordine teorico, ma rimandino a precise domande che la scuola si è posta in questi ultimi anni.

Uno degli obiettivi collegiali, nell'ambito dell'autonomia progettuale dell'attività educativa e didattica, è proprio il miglioramento della qualità della valutazione. In una recente circolare (C.M. n. 491 del 7 agosto 1996) che regolamenta la *Valutazione degli alunni della scuola elementare e dell'istruzione secondaria di primo grado*, per meglio organizzare le attività di valutazione, si esplicita l'opportunità «che ogni singola istituzione scolastica attivi un processo di ricerca in merito sia ai diversi aspetti delle prove di verifica (attendibilità e validità) sia ai criteri di valutazione delle prove stesse e all'elaborazione del giudizio complessivo» (art. 3).

4.1.1. La validità

La validità di una misura consiste nel grado di corrispondenza tra quella misura (filiazione quantitativa o qualitativa) e ciò (fenomeno, processo, dato) che si vuole misurare. Sinteticamente Garretti nel 1937 definì valido un test, se effettivamente misura quel che si propone di misurare. Sebbene la definizione di validità sia estremamente semplice, altrettanto non si può dire del suo controllo.

Pensiamo a una prova di padronanza o a un test che vogliamo valutare una determinata competenza. Prima di procedere alla valutazione della competenza o a un livello di prestazione, dovremo avere un qualche costrutto teorico che ci permetta di definire e chiarire componenti, relazioni tra di esse e ampiezza di quella competenza. In base a tale costrutto possiamo isolare determinati aspetti da sottoporre a verifica. Questo processo di scelta è ovviamente guidato da esigenze operative, e quindi porta a individuare elementi osservabili in base ai quali giudicare sulla competenza indagata. È seguendo questo modo di procedere che nelle verifiche che sottoponiamo alcuni compiti come rappresentativi di un insieme di operazioni, rivolgiamo alcune domande perché pensiamo che arrivino alcune forme di ragionamento, proponiamo alcune prove che richiedono per la loro esecuzione determinate procedure. Ma se tali compiti, domande e prove attivassero principalmente altre diverse operazioni, ragionamenti e procedure, potremmo dire che esse sono misure valide?

3. Le remariche della validità e dell'affidabilità delle misure e delle diverse modalità di controllo sono estremamente complesse per essere affrontate qui in modo eccessivamente sintetico. Per tale motivo rimandiamo un eventuale approfondimento a un testo particolarmente analitico: Boncorri (1993), pp. 113-302.

La validità di una misura e successivamente di una valutazione è quindi strettamente correlata al tipo di strumento che si adopera. Non è solo un problema di obiettivi, cioè su che cosa verta la prova, a richiedere il controllo della validità, ma anche la scala di misura che essa propone. Se adoperiamo o costruiamo uno strumento poco adatto alla misurazione di un determinato apprendimento o se utilizziamo una scala poco sensibile alla misurazione di quell'apprendimento, le nostre misure non saranno valide.

La validità degli strumenti di misura può essere verificata relativamente a contenuti, criteri, costrutti e alla modalità di presentazione.

1. *Validità di contenuto.* Se vogliamo valutare gli obiettivi raggiunti o un livello di competenza in un determinato settore, come per esempio nel caso di prove disciplinari, dobbiamo coprire il più ampio spettro di quel settore. Tale operazione, in alcuni casi, oltre a essere particolarmente impegnativa, potrebbe comportare competenze professionali e tempi di troppo superiori alle possibilità di un docente. Di fronte al bivio tra l'abbandonare il piano della verifica e il farne un'improvvisata, possiamo provare l'alternativa di "campionare" alcuni elementi del settore da sottoporre a verifica. Potremmo svolgere questa operazione da soli, o, preferibilmente, raggiungere un accordo tra esperti di quel settore o competenza. La costruzione e revisione della prova di verifica poggerebbero su quell'accordo e sulle competenze professionali del gruppo. Come decidere altrimenti sulla pertinenza di determinati aspetti o sulla quantità di alcuni di essi rispetto ad altri? Se inoltre vogliamo analizzare oungere a misurazioni di abilità trasversali o di competenze abbiamo bisogno di esplicitare il rapporto tra determinati comportamenti osservabili e un quadro teorico generale di quelle abilità o competenze. Nel misurare una particolare abilità linguistica, quale la comprensione della lettura, dovremmo esplicitare in via preliminare quali livelli di comprensione stiamo indagando o costruito teorico di riferimento stiamo adottando. Se volessimo utilizzare una prova di verifica di tipo strutturato, il modello ci aiuterebbe a costruire i quesiti che la compongono. Nel caso di prove meno strutturate permetterebbe di individuare gli stimoli più coerenti e mirati. Se il settore delle conoscenze o gli elementi/livelli di una determinata competenza fossero più ampi, il gruppo di esperti dovrebbe procedere a una fase di campionamento di quei contenuti e/o livelli: la prova, in questo caso, dovrebbe sondare una porzione significativa dei settori o delle competenze oggetto di studio. Il controllo della validità di contenuto in campo educativo si esplicita quindi attraverso:

- il riscontro della significatività degli elementi da sottoporre a verifica. Essi dovrebbero essere sostanzialmente centrali e utili per la valutazione;
- la scelta intenzionale e programmata (a campione, per quote proporzionali o con altri criteri esplicitabili) degli elementi da sottoporre a verifica come rappresentativi del settore di contenuti o della competenza da valutare.

2. *Validità di criterio.* Per verificare che cosa misura una determinata prova, si potrebbero utilizzare strumenti analoghi o misure diverse come misura-criterio. Il confronto tra risultati a diversi tipi di prove è quindi una verifica indiretta e si deve presupporre che lo strumento o misura che si adopera come misura-criterio sia valido a sua volta. Se per esempio usassimo i voti degli insegnanti di italiano per confrontare i risultati di un test di comprensione della lettura, potremmo scoprire una correlazione non significativa o in quanto i voti sintetizzano troppe competenze, oppure perché i voti non sono una misura valida. Il controllo della validità di criterio è possibile se disponiamo di misure simultanee, o fatte in un breve periodo. In questo caso parliamo di *validità concorrente (concurrent validity)*. Ma la misura-criterio può avvenire anche a distanza di tempo, e in questo caso parliamo di *validità predittiva*. Se per esempio l'ingresso in un determinato settore lavorativo fosse assunto come criterio di successo di un determinato corso di studio universitario, si avrebbe una misura di validità di quel corso solo a distanza di tempo.

3. *Validità di costrutto.* Nel caso in cui le nostre misure facessero riferimento a un costrutto teorico, potremmo a posteriori procedere a un'analisi con i risultati emersi da altre misure già note che fanno riferimento allo stesso costrutto (*validità convergente*). Analogamente potremmo verificare se misure che fanno riferimento a costrutti diversi fanno registrare basse correlazioni o risultano indipendenti (*validità discriminante*). La validità di costrutto è ovviamente riscontrabile solo se disponiamo di una teoria di riferimento che aiuti a individuare le componenti del fenomeno/comportamento oggetto di valutazione e quegli indicatori/descrittori che coerentemente permettono una sua misurazione.

4. *Validità di presentazione (o apparente).* Una prova o una misurazione che non tenga conto di alcuni elementi di impatto con chi vi si sottoporrà rischia di presentare elementi di distorsione. A volte è la presentazione grafica, troppo vistosa o eccessiva in lunghezza, a volte la contestualizzazione culturale o sociale a generare ambiguità, a volte la presenza di alcune formulazioni linguistiche troppo caratterizzate. Gli esempi potrebbero moltiplicarsi. Con la validazione di presentazione o di *faccetta (face validity)* si introduce il controllo di quegli aspetti qualitativi e intuitivi relativamente all'adeguatezza degli stimoli rispetto ai destinatari previsti.

4.1.2. L'affidabilità

L'affidabilità di una misura consiste nel grado di precisione con cui essa può essere compiuta. Il termine inglese *reliable* significa "che dà affidamento". Per richiamare il concetto di precisione, sebbene con leggere sfumature, si adoperano anche altre espressioni, tra le quali quelle di *fe-deltà, costanza, attendibilità*.

Lo strumento di misura, e in campo educativo possiamo pensare a una prova/compiuto strutturato quale un test, a una prova pratica o a un colloquio, per essere affidabile dovrebbe permettere una uniforme se non proprio identica rilevazione delle conoscenze e competenze da parte di diversi correttori/valutatori. Quanto più, rispetto a una stessa prova, la rilevazione di diversi correttori/valutatori è simile, tanto più possiamo assumere che lo strumento è affidabile.

In pratica l'affidabilità di uno strumento consiste nel permettere e garantire un basso livello di interpretabilità e quindi di uso soggettivo da parte di più persone. Diverse rilevazioni fatte dalla stessa persona o una stessa rilevazione fatta da più persone dovrebbero portare a risultati simili se non uguali. E questo sembra abbastanza intuitivo e banale se pensiamo a strumenti di misura rigorosamente precisi, quali per esempio le domande di un test. Se invece ci allontaniamo da strumenti di natura oggettiva, ci imbattiamo in problemi legati all'accordo tra più valutatori nell'uso di criteri, o con altre modalità di correzione/rilevazione. Il criterio dell'affidabilità è particolarmente sentito nell'uso dei descrittori o dei criteri guida per l'osservazione/misurazione di prestazioni.

A ben vedere l'affidabilità di una prova può essere considerata un aspetto della sua validità, in quanto se la prova non offrisse precise rilevazioni non ci permetterebbe di sostenere la sua validità complessiva. Ben dice Boncori (1993, p. 196) a riguardo: «Se una misurazione non è attendibile, è inutile indagare su altri aspetti della sua validità. Sarebbe come chiedersi qual è la forma geometrica dell'ameba o costruire castelli di sabbia su una spiaggia battuta dal mare mosso».

La precisione nelle misurazioni educative dipende dai diversi fattori che isolatamente o in modo concomitante si presentano nelle verifiche. Essi riguardano lo strumento di misura, il valutatore e il soggetto esaminato. La *precisione dello strumento* dipende da quanto riesce a offrire uno stimolo uguale per tutti, chiaro, e da quanto permette una registrazione dei risultati non soggettiva. Queste condizioni si verificano esclusivamente con le prove strutturate. Diversamente invece accade con altre forme di verifica. Quando le domande sono individualizzate come in un colloquio, o affrontate in gruppo, oppure si chiede di risolvere prove con tempi diversificati abbiamo bisogno di forme di valutazione che tengano conto delle diverse condizioni. Il controllo della precisione in questi casi è dato dal riscontro tra singole istruzioni, modalità dello stimolo e tipologia di risultati. Abbiamo bisogno di criteri da utilizzare individualmente o collegialmente e quindi l'affidabilità va costruita attraverso l'accordo tra correttori o rispetto ai singoli criteri definiti.

La *precisione del valutatore* dipende invece sia dalla variabilità nell'uso che può fare di uno strumento di rilevazione, sia dal grado di soggettività che può fare ricadere nell'attribuzione della misura. L'esamina-