

I dogmi della valutazione

Nel suo [articolo sulla valutazione](#) Vittoria Gallina ha esposto in maniera narrativa e convincente, cosa intendere per valutazione formativa, e ha lasciato giustamente aperto il problema del rapporto tra valutazioni dell'insegnante e valutazioni a livello di sistema. In effetti questo rapporto resta ancora problematico, non c'è stata nessuna 'composizione'. La valutazione dell'insegnante, non solo quella formativa ma anche quella sommativa (o conclusiva), che non può mancare, è *contestualizzata*, nel senso che tiene sempre conto di un percorso, con le sue caratteristiche specifiche, mentre quella standardizzata, di sistema, è per definizione *decontestualizzata*, in quanto tiene conto solo del risultato in prove uguali per tutti.

Bisognerebbe intanto chiedersi se la generalità dei risultati, ai quali tutti dovrebbero pervenire, debba necessariamente tradursi nella uniformità delle prove, le stesse per tutti. Si potrebbe porre questo problema come si pone, nella ricerca sociale, il rapporto tra la definizione concettuale di ciò che si indaga, il fenomeno che ci interessa conoscere (per esempio l'integrazione), e la sua operazionalizzazione, che è sempre relativa e ammette soluzioni diverse. L'uniformità delle prove, le stesse per tutti, un dogma del *testing* classico, potrebbe essere superata, pur salvaguardando la confrontabilità, mediante le diverse pratiche di equalizzazione (*equating*) che sono state elaborate.

Altro dogma è naturalmente quello dell'oggettività della valutazione, che richiede l'esclusione di ogni interpretazione (in quanto sempre soggettiva) dei risultati ottenuti. Le risposte al test non hanno bisogno di interpretazione, sono solo oggetto di elaborazione quantitativa; ma i dati che si ottengono, proprio perché non interpretabili, sono 'poveri' non ci dicono nulla del processo che ha portato alla risposta, e quindi della *qualità* della risposta. Si può ovviare, come a volte si fa, chiedendo allo studente di motivare la risposta data, ma c'è un limite: la motivazione, se diventa un resoconto sul ragionamento fatto, rende necessaria l'interpretazione, almeno in funzione di una codifica; e certamente non è una procedura adatta ai grandi numeri.

Come far emergere la qualità delle risposte

Il problema è quanto ci interessa capire la *qualità* delle risposte date, dei compiti svolti, dal punto di vista cognitivo. Se ci interessa, il test non deve soddisfare solo i criteri di validità rispetto al contenuto (la *content validity*), nel senso che dà una "rappresentazione" adeguata delle aree del curriculum e dei compiti scolastici; deve avere anche validità concettuale (*construct validity*), nel senso che, se con un quesito vogliamo misurare una determinata capacità, per esempio di *problem solving*, bisogna che sia questa l'attività cognitiva che rileviamo, e non altra^[1]. Verificare questo è difficile, perché una cosa sono le capacità *attribuite*, cioè ritenute essenziali per svolgimento di un determinato compito, un'altra l'attività cognitiva effettivamente svolta dallo studente.

Una verifica interessante di quanto sia attendibile la valutazione offerta dal test si può fare chiedendo a studenti, in colloqui individuali, di riflettere sulle risposte date e di motivarle, come è stato fatto, per esempio, da Campione et al. nel 1989 nel quadro di una ricerca sull'apprendimento della matematica a livello di scuola elementare^[2]. Sono stati intervistati bambini di classi IV e VI che avevano appena risposto a un test con quesiti a scelta multipla. Sulla base dei colloqui i bambini sono stati divisi in quattro gruppi, così caratterizzati: (a) risposte corrette e capacità di motivarle, (b) risposte corrette ma senza consapevolezza, (c) risposte errate senza capire perché, (d) risposte errate ma capacità di rendersi conto. In sostanza, considerando i risultati dei colloqui le valutazioni cambiavano. Le valutazioni offerte dal test si rivelavano errate nel 41% dei casi.

C'è poi un'altra questione, che fa entrare in gioco le scelte didattiche. Il test standardizzato, uguale per tutti, non può tenere conto dei percorsi nella loro specificità. Ammettiamo pure che a conclusione di un percorso scolastico conti solo il risultato osservabile, cioè la performance nelle prove. Ma se siamo ancora *nel* percorso, e magari all'inizio di un percorso scolastico, possiamo non tenerne conto?

Prendiamo il 'quesito del bus'^[3] (da un progetto di qualche anno fa), che presenta un tipico problema di scuola elementare. I dati del problema sono i seguenti:

- le tariffe della compagnia dei bus sono: una corsa 1 \$, la tessera settimanale 9 \$
- Yvonne prende l'autobus lunedì, mercoledì e venerdì per andare e tornare dal lavoro
- Anche martedì e giovedì lo prende per andare al lavoro ma torna a casa in macchina con gli amici

La domanda è: Yvonne risparmia se acquista la tessera settimanale?

Si chiede non solo di rispondere Sì o No ma di motivare la risposta. La risposta esatta è No, e la motivazione è data semplicemente riferendosi al numero di corse (che sono 8, con spesa 8\$). Ma vediamo come motiva la sua scelta un bambino che ha dato la risposta errata (Sì):

«Yvonne prende l'autobus 8 volte nella settimana, quindi spende 8 \$, ma se lo prende anche nel week end per andare a fare spese spende altri 2 \$, quindi le conviene avere la tessera».

Il bambino ha sbagliato la risposta perché ha riformulato il compito rendendolo più complesso. Ha sbagliato perché non ha risposto nei limiti posti dalla domanda, ma ha dato prova di capacità riflessiva (o immaginativa). Poi certamente imparerà questa regola linguistica, così importante nella nostra cultura, che bisogna rispondere solo nei termini posti dalla domanda, tenere conto quindi solo dei dati offerti dal problema (le ipotesi). Ma intanto viene penalizzato dal test che non tiene conto del fatto che siamo in un processo in corso, con aspetti interessanti da valutare.

Ma c'è qualcosa di più da dire: forse la riflessione sul problema, che ha portato il bambino a sbagliare risposta, dipende dal tipo di lavoro didattico, dal metodo seguito dall'insegnante che privilegia la discussione sui problemi, l'introduzione di varianti, la costruzione di nuovi problemi, metodo assolutamente apprezzabile, in linea con una concezione costruttivista della matematica. Il test dovrebbe tenerne conto, con quesiti differenziati, nel senso già detto. Ma questo significa anche che l'elaborazione dei quesiti dovrebbe presupporre una conoscenza approfondita delle alternative pedagogico-didattiche.

L'errore di 'lavorare per il test'

Lavorare per il test, cioè ridurre l'attività didattica alla preparazione in vista del *test behavior* che sarà richiesto, è una scelta deprecabile, perché isola certi risultati, che sono validi solo in quanto elementi di uno sviluppo intellettuale complessivo, come fini in sé. Bisogna evitare che siano penalizzati proprio i percorsi didattici più interessanti, *process-oriented*, che magari tardano a raggiungere risultati 'spendibili'.

Prove che spesso contrastano con una ragionevole impostazione didattica sono per esempio, sempre a livello elementare, quelle di lettura e comprensione, che richiedono, per poter rispondere alle domande, una lettura del testo estremamente analitica, capace di estrarne tutti gli elementi. Ma nessuno (bambino o adulto) legge così; 'ragionevole' è la lettura selettiva, che si fa in base a qualche interesse, problema, o attività. Se questo è vero bisogna dire che se ci si vuole 'preparare al test' si rischia di prendere una strada didatticamente errata, quella dell'esercizio fine a se stesso.

Comunque, a proposito del rapporto tra valutazione da parte degli insegnanti e valutazione standardizzata a livello di sistema, una fondamentale osservazione da fare è che nessuna prova è esaustiva, tutte sono relative. Poi, per caratterizzare il possibile rapporto, è necessario avere dati che siano frutto di sperimentazioni mirate. Al momento ci mancano anche dati elementari su cui riflettere, come le correlazioni, magari per un campione di classi, tra le valutazioni degli insegnanti e i punteggi nel test.

[1] l'importanza della *construct validity* è stata sostenuta da Lee Cronbach fin dagli anni '50 in alternativa alla concezione dominante, tutta metrologica, del testing.

[2] J.C. Campione et Al. (1989), *Metacognition*, in R.I.Charles, E.A. Silver (eds), *The teaching and assessing of mathematical problem solving*, NCTM.

[3] Dal progetto Quasar, in M. Niss (1993), *Cases of assessment in mathematics education*, Kluwer Academic, Dordrecht

Maurizio Lichtner Esperto di ricerca educativa, ha insegnato Pedagogia alla Facoltà di Sociologia della Sapienza ed Educazione degli adulti all'Università di Roma Tre. Si è sempre occupato di valutazione e sul tema ha pubblicato nel 2003 *Valutare l'apprendimento: teorie e metodi*, Franco Angeli.