

Dal pasticcio pilota al pasticcio generale

[Proteofaresapere](http://www.proteofaresapere.it)

18-05-2005

E se qualcuno valutasse i valutatori...

di *Piero Lucisano*

La perseveranza con cui, dal progetto pilota 1 ai progetti pilota 2 e 3 all'attuale pomposissimo Servizio Nazionale per la Valutazione attuale, l'Invalsi riformato sotto la guida della coppia Aprea Elias continua a riproporre strumenti di rilevazione non validi e non affidabili, costruiti in modo approssimativo, somministrati con procedure caserecce e analizzati in modo superficiale, ha del sublime. E' difficile infatti trovare le parole per esprimere i sentimenti che provoca un abisso che al tempo stesso atterrisce e attrae.

L'idea è che il Sistema Nazionale per la Valutazione debba basarsi su test direttamente somministrati a tutti gli studenti del paese. A questi test si dovranno in sostanza uniformare tutte le scuole e la Aprea vagheggia anche un manuale di preparazione che aiuti gli insegnanti e gli studenti ad adeguarsi agli standard previsti dal governo. In questo modo le scuole potranno agire la loro autonomia, adeguandosi ai dettami del Ministero dell'Istruzione.

Nel condurre i progetti pilota uno due e tre, l'ingegner Elias ha evidenziato come sia difficile il transfer di apprendimenti; infatti le metodologie che insegna sulle misure ortofrutticole, nonché il piglio ingegneristico lo conducono a mostrare una efficienza pari solo alla quantità di errori che è riuscito ad accumulare in una sola attività di ricerca e a ripetere gli stessi errori in tre progetti pilota consecutivi ed infine riprodurli senza variazioni anche nella somministrazione obbligatoria delle prove in tutte le scuole elementari e medie (senza pietà né per gli studenti, né per le scuole, né per le tasche dei contribuenti).

Le prove sono state costruite da un Gruppo di lavoro diretto da Elias stesso in poche settimane. Le prove sono straordinarie perché senza alcun costrutto teorico riescono, a detta dei protagonisti, a misurare con poche domande chiuse competenze complesse quali l'italiano, le scienze, la matematica. Per ciascuna di queste discipline e per ogni livello scolastico, l'ingegnere ed il suo gruppo sono riusciti a individuare una trentina di domande che assolvono all'intera descrizione delle competenze in questi ambiti. In termini tecnici queste osservazioni si riferiscono alla validità di contenuto. Il concetto di validità viene utilizzato per verificare se una prova misura ciò che intende misurare. Riferito ai contenuti significa che, definito ciò che si intende misurare, si tratta di costruire un numero di prove/item tali da coprire in modo ragionevole l'area di contenuti da esaminare.

Nelle indagini internazionali si utilizzano per verificare settori specifici di competenze come, ad esempio, le capacità di base di comprensione della lettura, normalmente più di cento domande; quando si vuole verificare la competenza in Italiano il problema è più complesso. In una indagine sulla scuola elementare svolta dal CENSIS per lo stesso Ministero ad esempio per approssimare una valutazione sulle abilità in lingua italiana erano previste oltre cento domande e due prove di produzione scritta. Nelle ultime ricerche sia la IEA, sia l'OCSE utilizzano inoltre domande aperte perché non si accontentano delle risposte chiuse e investono considerevoli risorse per analizzarne i risultati.

Un'ulteriore osservazione riguarda l'affidabilità delle prove. Tradotto si tratta di vedere se le prove predisposte misurano in modo preciso. L'affidabilità delle prove si consegue attraverso un complesso processo di messa a punto che richiede la sperimentazione delle stesse prove su campioni omogenei a quelli su cui si intende utilizzare la prova. Questo richiede che si programmi per tempo e nelle indagini serie il try out si realizza un anno prima della somministrazione principale.

Invece nella rilevazione nazionale del sistema istruzione non si perde tempo: le prove vengono messe a punto e tarate in due mesi, da ottobre a metà novembre, dunque su studenti di età diversa e in periodi dell'anno diversi ("somministrati a metà novembre in una classe successiva a quella oggetto di rilevazione"). Si fa presto, qualche classe di insegnanti amici e poi basta. Vengono poi utilizzati i modelli statistici più avanzati per verificare gli esiti di questa operazione, ma le decisioni vengono assunte prescindendo dalle regole che la comunità scientifica si è data per la messa a punto, la validazione e la taratura di un test. L'Invalsi dichiara ad esempio di utilizzare come criterio di selezione degli item la sola facilità delle domande. I dati di queste procedure non sono pubblicati e se ne danno solo accenni sommari.

Del resto dopo che i dati del Progetto pilota Due erano stati sottoposti a critica, l'Invalsi ha ritenuto di non esporsi, evitando del tutto la pubblicazione dei dati del progetto pilota tre.

Pensate che i docimologi sostengono che per tarare una domanda è necessario che si verifichi il suo indice di discriminatività e i più pignoli chiedono che se ne esamini la fitness sulla base di calcoli complessi richiesti dalla cosiddetta Item Response Theory. I docimologi pensano, inoltre, che le domande di una prova per misurare debbono avere diversi valori di difficoltà altrimenti, dicono, si rischia di non cogliere le differenze tra le prestazioni migliori e quelle medie.

Ma l'Invalsi in tre anni di progetti pilota non è riuscito a programmare un ciclo corretto: messa a punto, taratura delle prove somministrazione principale. E non ha tempo da perdere in sottigliezza. Pensate che i docimologi, quelli seri, si mettono a sottilizzare anche sui testi da usare per misurare la comprensione della lettura. Non avrebbero mai usato domande quali le seguenti, più adatte alla settimana enigmistica o a giochi di società che a rilevare competenze puntuali. Come si può osservare, infatti, in tutte e tre le domande ciò che è rilevante è l'abilità dello studente nel comprendere la modalità fuorviante con cui è proposto il quesito.

Su una scatola di spaghetti c'è scritto

Spaghetti

Cottura 12 minuti 500 grammi

Per cuocere, secondo le indicazioni 250 g. di spaghetti quanti minuti sono necessari?

- A) 9
- B) 24
- C) 12
- D) 6

Quanto può pesare un uovo di gallina?

- A) 250 g
- B) 50 g
- C) 5 mg
- D) 1,5 hg

La parte decimale dei fattori della moltiplicazione è stata coperta

8, // // 25, // //

quale può essere il risultato corretto?

- A) 2,11328
- B) 2113,28
- C) 211,328
- D) 21,1328

Pensate che questi ricercatori noiosissimi cercano di stabilire il tempo utile alla somministrazione delle prove con molta attenzione e invece noi di tempo ne abbiamo dato a volontà, tanto a scuola non fanno nulla di utile.

Inoltre c'è stato chi ha notato che i progetti pilota hanno avuto risultati diversi da tutte le ricerche precedenti: il centro nord ottiene risultati inferiori al sud. Per quanto riguarda il progetto pilota tre la cosa non si è potuta verificare perché i risultati non sono stati pubblicati. Questo è di nuovo un problema di validità. Si tratta di quella che viene definita validità del criterio: misurando la stessa cosa con strumenti diversi si assume che si dovrebbe avere all'incirca lo stesso risultato. Ovvero o il modello di rilevazione di alcune decine di indagini nazionali promosse dallo IEA, dall'OCSE e dallo stesso Ministero sono sbagliate o è sbagliato il modello applicato dal Servizio Nazionale di valutazione dato che misurando lo stesso fenomeno hanno esiti contrapposti.

Tuttavia chi volesse informarsi sui risultati potrebbe leggere dal sito Invalsi "Le prove della seconda classe della scuola elementare sono andate bene, i bambini hanno potuto rispondere alle domande senza difficoltà" manca solo che scrivano che se si fossero impegnati di più avrebbero potuto fare meglio e si sarebbe raggiunto l'apice delle banalità educative.

Ovviamente tutto si può spiegare, ad esempio per l'italiano si potrebbe ipotizzare che in alcune aree del centro nord siccome la segnaletica è riportata in dialetto hanno qualche problema in più. Per la matematica poi il problema potrebbe avere origine nella sempre maggiore diffidenza per i numeri arabi. Noi lo sapevamo già e bisogna sostenere il Centro nord se va peggio e premiare il centro nord se va meglio. In sostanza tutti i problemi vengono dalla presenza del centro sud. Basterebbe eliminare i dati del centro

sud per non avere più questo problema.

Alcuni critici hanno poi detto che senza questionari di sfondo la ricerca è cieca perché non ha senso rilevare misure senza cercare di capirne le ragioni. Ma siccome le ragioni le conosce già il Ministero, è inutile sprecare tempo e risorse. E poi, anche se ci sono ragioni che non conosce, l'importante è dare le risposte agli insegnanti in tempo reale. Il progetto Pilota Due a giugno aveva restituito risultati a un milione di studenti.

Certo ci sono dei pignoli che dicono che i risultati riportati nel rapporto, oltre agli errori nella scelta delle prove, nella taratura delle domande, nei confronti di criterio, presentino anche vistosi errori di calcolo. "Stiamo a guardare il capello!" Numero più numero meno, quello che conta è che la gente pensi che siano corretti, tanto poi le decisioni le prende il governo e non bisogna criticarle.

Nella rilevazione del Servizio Nazionale di Valutazione non sono previsti item di ancoraggio tra i diversi ordini di scuola. Del resto il presidente dell'INVALSI ritiene che si possano confrontare i risultati di ordini di scuole diverse su prove diverse sulla base delle percentuali di risposte corrette e tanto basti.

Ai progetti pilota le scuole hanno dovuto partecipare tutte volontariamente. Chi non si fosse proposto volontario avrebbe dimostrato di essere poco volenteroso. Ora l'obbligo di partecipare alle prove è di legge. Adesso abbiamo assistito ai primi momenti di ribellione da parte delle scuole e delle famiglie ed il rischio è che il lavoro dell'Invalsi finisca per convincere che l'acqua è così sporca che tanto vale buttare pure il bambino.

La valutazione è una cosa seria, come è serio il lavoro di mettere a punto prove, di valutarne gli esiti, di restituirle al sistema scolastico perché le utilizzi per autovalutarsi. Il sistema nazionale di valutazione centralizzato e dilettantistico impedisce che si consolidi quel corretto approccio alla valutazione che nel nostro paese ha richiesto anni di sforzo di insegnanti e ricercatori per essere costruito.

Un serio sistema nazionale di valutazione dovrebbe concentrare i suoi sforzi nella predisposizione di prove valide ed affidabili su tutti gli ambiti disciplinari previsti dai curricula. Questo si può fare in maniera seria su un campione di scuole, con modalità di somministrazione rigorose, garantendo alle stesse scuole l'uso riservato dei risultati. Con un ritmo biennale si dovrebbero mettere a punto e tarare prove su due forme su campioni casuali nazionali, e rilasciare di volta in volta una delle due forme del test alle scuole con le indicazioni per utilizzarla per la autovalutazione, conservando invece una forma della prova per la successiva somministrazione nazionale. Questa procedura garantirebbe la qualità dei dati e la possibilità di svolgere comparazioni diacroniche. Le prove per ogni ordine di scuola dovrebbero contenere item di ancoraggio per garantire i confronti tra i diversi cicli scolastici. Inoltre, sarebbe importante che nelle rilevazioni campionarie si facesse un uso attento dei questionari utili a comprendere le ragioni prevalenti dei risultati osservati e dunque per verificare quali siano gli aspetti di efficacia e quali i punti deboli sia dei sistemi di insegnamento, sia del sistema formativo nel suo complesso.

Quello che è straordinario è che queste indicazioni porterebbero alla scuola una grande ricchezza di strumenti, un approccio più corretto alla valutazione e che il tutto costerebbe meno della metà di quello che si spreca con l'attuale sistema.