

VALUTAZIONE DELLE SCUOLE E VALUTAZIONE DEGLI APPRENDIMENTI.

Nietzsche dice (parafraso) “il buon maestro è quello che insegna all’allievo a tradirlo...”. A parte ogni apprezzamento all’aforisma che concentra troppe considerazioni analitiche per essere qui esplorato compiutamente (per esempio “fulmina” ogni pericolo e deriva “pigmalionica” che è sempre in agguato, come sa ogni buon docente) il concetto espresso dal filosofo contiene un principio di fondo che presiede la filosofia e la pratica della “formazione”.

Essa si sottrae ampiamente, anzi nei suoi fondamenti, al paradigma del “risultato”.

Nulla di più lontano da tale paradigma infatti del “tradimento”.

O, in altre, forse per alcuni, più accettabili termini: la formazione “forma hominis juxta propria principia (Tommaso)” è l’acquisizione progressiva dell’esercizio della autonomia della persona. E dunque si sottrae “in fondamento filosofico” alla misura del risultato.

Ma quel processo (la formazione) accade nella modernità entro un “contenitore” organizzato (il, i, sistemi di istruzione nazionale) che ne traduce il “valore”, in termini incondizionati, reso disponibile (progressivamente) a tutti i cittadini (l’universo delle nuove generazioni e non solo).

D’altra parte, almeno a partire dalla seconda rivoluzione industriale, quel “valore assoluto” (il sapere, la formazione come diritto di tutti) si incorpora più o meno direttamente nello sviluppo economico. Dunque, assume anche un “valore condizionato” (economico, appunto).

In termini marxiani, il primo (il valore incondizionato) è “valore d’uso” proprio dell’uomo. Il secondo è “valore di scambio”, condizionato dai rapporti di produzione e di scambio. Dal mercato insomma.

Ma lo sviluppo dei sistemi di istruzione e formazione a partire da quella fase storica (culmine nella seconda metà del secolo scorso), non fosse altro che per le risorse economiche, organizzative, professionali e per il coinvolgimento di masse ingenti sia di discepoli che di “maestri” e per il suo rapporto con lo sviluppo economico complessivo, pone il problema della “misura” e del “risultato”. Con essi si raffrontano le “convenienze sociali” (il finanziamento attraverso la fiscalità), gli imperativi etici e politici (usare al meglio le risorse comuni) e il rapporto, certo non meccanicistico, con lo sviluppo economico.

La “dialettica” della formazione è dunque compresa entro il reciproco dipanarsi e confrontarsi dei “due valori” indicati: quello incondizionato, non esplorabile con il paradigma del risultato, e quello “condizionato”, economico, vincolato al paradigma stesso. Tale dialettica rappresenta l’impegno, la fatica (e spesso il dolore) di chi si occupa di formazione.

E’ dunque assolutamente legittimo e necessario che, per un macroaggregato sociale ed economico, come quello dedicato alla “produzione” del diritto all’istruzione ed al sapere per tutti, si pongano interrogativi di valutazione, economica e non, del rapporto tra risorse impegnate (intese in senso lato: economiche, professionali, culturali, di senso comune...) e risultati dei processi stessi.

Ma per produrre elementi significativi tale necessità (la valutazione dei risultati) deve saper misurarsi con la dialettica citata, specificando con accuratezza il “cosa”, il “chi” e il “come” e il “perché”, e i limiti e le condizioni di una assennata valutazione. Innanzi tutto riconoscendo la “costitutiva” parzialità insita nell’assumere i “risultati di apprendimento” come “esaustivi” del paradigma stesso di “risultato”.

Ai sistemi di istruzione nazionale, nel loro consistere storico come sub articolazioni essenziali dell’assetto istituzionale, sociale, economico di una formazione sociale, sono infatti assegnati obiettivi e funzioni che, pur facendo dell’apprendimento la missione fondamentale, ad essa non si limitano.

Basti pensare a funzioni essenziali come quelle rappresentate dalla “equità sociale”, dal contributo alla colmatatura delle differenze, alla promozione di “socializzazione” culturale, etica, comportamentale, ed all’insieme dei “servizi” che le scuole sono chiamate ad organizzare per risponde a tali missioni.

A livello internazionale tale preoccupazione si esercita (e da anni) e contrassegna i caratteri dei diversi sistemi di valutazione dei sistemi di istruzione.

Senza approfondire una comparata assai varia, si possono rilevare elementi essenziali

1. La valutazione degli apprendimenti è sempre affidata ai docenti ed alla loro “cultura professionale” che va costantemente aggiornata e sviluppata.
2. Il “sistema” si cimenta con “rilevazioni standard” effettuate su alcuni apprendimenti. In sostanza lingua madre e straniera, matematica, TIC. Per consapevolezza più che fondata non si sottopongono a rilevazioni standard altri apprendimenti per ragioni relative alla “ampiezza epistemologica” di molte delle discipline dei insegnamento (da quelle storico-filosofiche a quelle che compongono le “scienze naturali”) o meglio sulla indeterminatezza della loro traduzione in chiave di curricolo di istruzione.
3. Le rilevazioni possono essere sull’universo o su un assennato campione. I loro risultati non sostituiscono la “valutazione dei docenti” ma si offrono come “indicatori” sia a livello micro (i docenti delle singole scuole) sia a livello macro (i decisori di sistema).
I primi, dalla analisi dei risultati intesi come “indicatori” dei processi a livello di scuola, possono ricavare utili feedback del loro stesso lavoro, individuando obiettivi di rinforzo, di correzione, di miglioramento, attraverso un esame dettagliato ed analitico
I secondi possono ricavare dai dati aggregati e confrontati a livello di sistema, feedback delle politiche pubbliche seguite, indicazioni di correzioni necessarie e di politiche di investimento e sviluppo.
4. Le politiche pubbliche che utilizzano i processi valutativi come feedback e controllo della “razionalità decisoria” della politica stessa possono operare secondo strategie diverse, più o meno condivisibili politicamente (p.es. con un più che discutibile legame tra esiti valutativi degli apprendimenti e politiche finanziarie). Ma in ogni caso viene reso esplicito (dunque sottoposto a processi di controllo e consenso) tale rapporto, a partire dallo stesso “riconoscimento sociale” del valore e della “reputazione” del sistema e dei protocolli valutativi.
Da tale punto di vista nel nostro Paese siamo agli inizi sia della costruzione del sistema, sia della costruzione di una necessaria “cultura sociale” e riconoscimento reputazionale, in assenza dei quali il sistema valutativo stesso semplicemente non funziona
5. Le rilevazioni standard, che consentono la comparazione e l’analisi all’interno di un macrosistema di necessità composito e differenziato, si prestano ovviamente a limiti di significato che sono attentamente esplorati. Fenomeni di cheating, di opportunismo nei comportamenti, impropri cortocircuiti tra “valutazione degli apprendimenti” e “rilevazioni standardizzate” (teaching to test) sono controllati sia per via di analisi statistiche (anche il solo esame analitico dei risultati rilevativi è in grado di rivelare cheating p. es.) sia in riferimento all’etica professionale” dei docenti stessi (controllo del fenomeno del taching to test per es.). Da qui la sottolineatura del valore essenziale che riveste il riconoscimento sociale e la reputazione condivisa del sistema di valutazione per il suo stesso “successo”.

Molta polemica corrente che contraddistingue la discussione nazionale sul “Sistema di valutazione” è alimentata (colpevolmente) da equivoci che ignorano le proposizioni precedenti: a partire dalla stessa confusione terminologica che indica come “valutazione” le “rilevazioni sui livelli di apprendimento” o che cortocircuita la “valutazione di risultato” sugli esiti di quelle rilevazioni, ignorandone la funzione di “indicatori” di uno dei compiti della scuola (sia pure il più importante, ma non l’unico).

Ciò che si perde (colpevolmente) in quegli equivoci è il diverso significato che devono assumere i “dati” delle rilevazioni sui livelli di apprendimento, quando siano strumento utile per il “decisore” della politica pubblica e quando, viceversa, siano materiale diagnostico per l’autoregolazione del “produttore” (della istituzione scolastica autonoma).

A livello micro (il singolo produttore) l’esame analitico degli esiti delle rilevazioni standard (esplorato sui livelli più disaggregati) è utile per rispecchiare, in feedback, l’effetto di impatto di scelte didattiche ed organizzative, e il livello di efficacia ed efficienza espresso nella propria programmazione (risorse, impegno del personale, progettazione, sviluppo organizzativo)

A livello macro (il decisore della politica pubblica) l'esame degli esiti delle rilevazioni standard (il cui livello analitico esplorato è evidentemente diverso dal precedente e può fermarsi a più ampi aggregati) può/deve costituire il primo step di un feedback di valutazione che dalle misure di impatto ed efficacia ed efficienza del singolo produttore risale alle scelte della stessa politica pubblica. (le scelte strategiche, le "teorie" che le ispirano, i programmi e le risorse di sistema messi in atto in conseguenza di esse).

Va ovviamente considerato fisiologico il fatto che i due livelli elaborino interessi e sensibilità non coincidenti, a volte contraddittori e addirittura conflittuali. Ma, posto che "la valutazione non è (mai) una festa", si tratta di gestire tale dialettica facendone uno strumento di crescita complessiva e non di blocco reciproco.

Sono evidenti, su entrambi i fronti, due polarizzazioni estreme.

Il decisore della politica pubblica limitando la valutazione al livello dell'impatto e dell'efficienza del singolo produttore, e non investendo la scala ascendente (programmi, risorse, strategie, "teorie" del "bene pubblico" a cui si ispirano) mostra di "scaricare" eventuali fallimenti o limiti nei risultati sull'inefficacia/inefficienza del singolo produttore. E tale rischio è tanto più accentuato quando la politica pubblica scelga la strada del rating, delle classifiche, del meccanicismo causale nel nesso risultati-risorse.

Sul fronte opposto il singolo produttore tende ad ascrivere i risultati non soddisfacenti ad un repertorio di "vincoli e condizioni esterne" la cui responsabilità viene totalmente ascritta al decisore pubblico (limiti nelle risorse, contesti sociali sfavorevoli, inconfondibilità dei dati, difetti nelle procedure rilevative ecc..). Fino a negare "per principio" lo stesso ruolo della valutazione, oppure conformando opportunisticamente i propri comportamenti. (Basta questo per "mandare fuori bersaglio" qualunque protocollo valutativo).

Certo vi è da aggiungere che tra i due possibili opportunismi, quello ascrivibile al decisore della politica pubblica è il più grave e comporta le più gravi conseguenze (la politica pubblica si sottrae allo strumento valutativo che essa stessa vorrebbe abilitare, e compromette così, ab origine, il diffondersi e consolidarsi di una "cultura della valutazione").

Se ci si pone a livello "micro", come interpretare al meglio la funzione che gli esiti delle rilevazioni sui livelli di apprendimento possono esercitare in termini di autoregolazione, autovalutazione, feedback sulle scelte di programmazione effettuate?

Più di molte argomentazioni si propone un semplicissimo esempio che esplora il livello assolutamente iniziale (e rudimentale) di tale lavoro. L'esempio è tale se serve sia ad indicare alcuni equivoci che alimentano incomprensioni e polemiche, sia le prospettive di approfondimenti necessari e possibili per raggiungere livelli di ben più consistente significato di quelli che dall'esempio stesso emergono.

Il materiale di esempio è costituito dagli esiti delle rilevazioni dei livelli di apprendimento in un istituto di secondaria superiore di una regione compresa nella ripartizione NordEst, su quattro classi seconde. I dati sono reali e ovviamente anonimizzati.

Il lettore potrà cimentarsi con le analisi più dettagliate. Ci si limita sottolineare gli elementi che si rendono più evidenti ad un osservatore esterno.

1. nel confronto sui punteggi medi e complessivi la scuola appare sostanzialmente allineata con il contesto regionale e di area geografica e collocata nella fascia alta rispetto alle medie nazionali, con qualche rilievo non parimenti confortante per quanto attiene alle rilevazioni in Italiano.
2. La distribuzione dei risultati tra le classi appalesa un evidente squilibrio: la "classe a" presenta dati non solo inferiori alla media della scuola ma anche a quelli degli altri riferimenti. Si noti che i risultati relativi alla classe sono collocati al di sotto del limite inferiore sia delle medie regionali, sia dell'ambito geografico. A parità di altri fattori (p.es. particolari discontinuità nel corpo docente) i dati rappresentano un sintomo probabile di non soddisfacente processo di formazione delle classi. Il risultato non è soddisfacente rispetto a obiettivi di equità interna.

3. La rilevazione complessivamente soddisfacente relativa alle altre tre classi mostra andamenti contraddittori in relazione ai diversi “ambiti” che, per ciascuna disciplina, articolano le prove INVALSI. Non sono dati “negativi” ma sottolineano differenze per esempio relative alla padronanza del “testo narrativo” o nel “testo espositivo” e in “grammatica” che non mantengono gli alti livelli medi (Classi c e d). Medesimi rilievi si possono fare per Matematica nell’ambito “spazi e figure” (classe c).
4. Se un osservatore esterno segnala ad una prima analisi alcune contraddizioni, il lavoro di esame analitico condotto dagli stessi docenti, con la conoscenza dettagliata delle singole condizioni operative, potrebbe non solo approfondire la diagnosi, ma altresì elaborare misure di miglioramento: per esempio la debolezza nella padronanza del “testo espositivo” potrebbe coinvolgere i docenti di scienze (elaborare una buona relazione di una esperienza di laboratorio è ottimo esercizio) o di lingua straniera. I docenti di Matematica insieme a quelli di disegno affrontare la relativa debolezza segnalata nell’ambito “spazio figure”. I docenti collettivamente proporsi di affrontare le disparità dei risultati tra classe e classe, anche in presenza di buoni punteggi medi, in termini di qualche misura di migliore programmazione comune disciplinare e transdisciplinare.
Dal punto di vista della programmazione e gestione complessiva “tutti” dovrebbero rielaborare più assennate metodologie nei criteri di formazione delle classi e nel controllo della loro applicazione.

Come si vede, anche da un approccio assolutamente esterno e superficiale, quale non può che svilupparsi in questa sede, emergono elementi di feedback interessanti: un gruppo di docenti che si misurasse sul campo con tale analisi ne può ricavare (anche “falsificando”, ma con impegno critico, i dati relativi) significativi input di riorganizzazione e miglioramento del comune lavoro.

L’utilità di simili “ritorni” provenienti dalle rilevazioni non ha bisogno di ulteriori argomentazioni. Il problema è costituito dal fatto che meno di un terzo delle scuole adotta un simile atteggiamento di approfondimento analitico dei dati “restituiti” dalle rilevazioni. In qualche caso (un altro terzo) ci si accontenta di una analisi che si ferma ai dati di scuola ed ai confronti sui “valori medi” (così contraddittoriamente avvalorando le semplificazioni che si vorrebbero combattere). In altri casi infine, i dati restituiti dall’INVALSI sono “a disposizione”, ma non diventano oggetto di tale confronto collettivo.

In tal modo non solo si “spreca una occasione”, ma contemporaneamente si compromette lo stesso impegno che sarebbe possibile e necessario per migliorare lo stesso strumento di rilevazione.

Gli strumenti della rilevazione, infatti, non sono “mai” i migliori possibili. Hanno un indubbio valore di fondo costituito dal fatto che tutta la popolazione scolastica si cimenti con il medesimo strumento e dunque producono dati confrontabili.

Ma la valutazione è, per definizione, sempre un campo di ricerca. Gli strumenti di rilevazione possono e debbono essere sempre migliorati. Gli “esperti” e i “tecnici” avrebbero sempre bisogno del riscontro critico di chi sta “sul campo”.

E, d’altra parte, se dovessimo attendere di avere “i migliori strumenti” prima di cimentarci con questa opera di ricerca, essa non comincerebbe mai (E’ la storia del sistema di valutazione nel nostro Paese..).

Sul fronte del decisore della politica pubblica queste analisi hanno significato?

Come si vedrà più avanti il decisore pubblico deve misurarsi con una “scala” di valutazione che è diversa. Ma anche a tale livello vi sono “inferenze” utili che dilatano la loro portata sul piano della politica pubblica.

Per esempio le notazioni qui condotte sul tema della “equità” del servizio reso ai cittadini (nel caso micro i criteri inadeguati di formazione delle classi) si dilatano a livello territoriale se si esaminano le differenze (più significative) dei dati delle rilevazioni, sul territorio nazionale.

A livello provinciale, regionale, di comparto geografico, come assicurare il livello di equità del servizio pubblico? Quali misure e scelte di politica pubblica per combattere le differenze? Quali

“differenziali” nella politiche di investimento (e non si tratta semplicemente di Nord e Sud: la questione investe il complesso della dimensione territoriale del servizio di istruzione). Ancora: alcune “debolezze” riscontrabili nei valori e nella distribuzione dei dati delle rilevazioni per “ambiti” all’interno di ciascuna disciplina possono essere, quando convalidate a livello generale, ispirazioni per politiche culturali o per iniziative generali di ricerca didattica. L’educazione linguistica p.es. non riguarda solamente i docenti di Italiano, ma di tutte le discipline. Debolezze estese nella padronanza del testo espositivo possono interessare l’intero uso “regolativo-normativo” della lingua e dunque, per esempio, le discipline scientifiche. La debolezza (confermata) negli apprendimenti di Matematica e in particolare nell’ambito della geometria possono suggerire iniziative di formazione che non riguardano solo i docenti di matematica. La capacità e padronanza della lettura di dati, della statistica, estende la propria influenza su discipline come la Storia, la Geografia, e così via. Il decisore pubblico potrebbe ricavarne indicazioni di sviluppo specifiche per attività di formazione, ricerca, sperimentazione. Una strategia pubblica capace di misurarsi con i riflessi che provengono dai dati delle rilevazioni degli apprendimenti, scongiurando il rischio di cortocircuitare su di essi ogni impegno valutativo, potrebbe selezionare e valorizzare anche scelte “organizzative” di fondo, relative a spazi e ambienti formativi, tempi di insegnamento, strumentazione didattica. Criteri di destinazione degli investimenti. Insomma mettendo anche in discussione, con il supporto dei dati oltre che delle “opinioni e teorie”, assetti fondamentali dell’organizzazione del sistema. Ma tutto ciò a due condizioni: assumere i dati delle rilevazioni dei livelli di apprendimento come “indicatori” o sintomi per rielaborare diagnosi (e non “classifiche e graduatorie”); dunque avendo ben presente che “valutazione dei risultati” è paradigma che va ben oltre alla semplice “valutazione dei livelli di apprendimento”. E, in secondo luogo, sottomettendosi alla logica della valutazione, e accettandone le inferenze fino al livello della determinazione della strategia pubblica. Si valuta “per decidere”, cioè per potenziare la “razionalità decisoria” da cui muove (o lo dovrebbe) il decisore politico e amministrativo. Insomma accettare ed anzi promuovere l’esplorazione dell’intera filiera della valutazione: da ciò che accade in ogni singola scuola, a ciò che accade in Viale Trastevere o a Palazzo Chigi.

Punteggi complessivi	Limite Inf	Italiano	Limite Sup	Limite Inf	Matematica	Limite Sup
Classe a	-	56,6	-	-	45,2	-
Classe b	-	66,8	-	-	57,0	-
Classe c	-	72,1	-	-	57,0	-
Classe d	-	70,6	-	-	60,3	-
Scuola	-	66,3	-	-	54,7	-
Regione	68,6	70,8	73,1	48,5	51,4	54,3
Nordest	72,0	72,9	73,8	50,6	52,3	54,0
Italia	69,3	69,8	70,3	47,2	47,8	48,4

Punteggi matematica	Limite Inf	Numeri	Limite Sup	Limite Inf	Spazio e figure	Limite Sup	Limite Inf	Dati e previsioni	Limite Sup	Limite Inf	Relazioni e funzioni	Limite Sup
Classe a	-	34,2	-	-	47,3	-	-	65,4	-	-	34,5	-
Classe b	-	46,8	-	-	66,6	-	-	76,1	-	-	40,4	-
Classe c	-	55,5	-	-	54,5	-	-	76,5	-	-	40,8	-
Classe d	-	55,4	-	-	59,8	-	-	84,0	-	-	42,0	-
scuola	-	47,7	-	-	56,9	-	-	75,2	-	-	39,3	-
Regione	38,7	42,5	46,3	49,1	52,4	55,7	71,3	73,7	76,2	34,9	37,3	39,7
Nordest	40,3	42,5	44,7	52,1	54,3	56,4	74,1	75,3	76,5	36,0	37,4	38,8
Italia	37,3	38,1	38,8	49,0	49,8	50,5	69,3	69,8	70,4	33,3	33,9	34,5

Punteggi Italiano	Limite Inf	Testo Narrativo	Limite Sup	Limite Inf	Testo Espositivo	Limite Sup	Limite Inf	Testo Argomentativo	Limite Sup	Limite Inf	Grammatica	Limite Sup
Classe a	-	61,9	-	-	68,4	-	-	58,4	-	-	45,6	-
Classe b	-	69,7	-	-	65,2	-	-	73,6	-	-	55,3	-
Classe c	-	70,5	-	-	81,9	-	-	81,4	-	-	60,8	-
Classe d	-	65,3	-	-	76,4	-	-	77,5	-	-	67,6	-
Scuola	-	66,8	-	-	72,9	-	-	72,5	-	-	57,0	-
Regione	64,8	67,3	69,7	76,8	79,3	81,8	73,8	75,6	77,3	64,8	67,8	70,8
Nordest	68,4	69,5	70,5	79,7	80,8	81,8	76,7	77,5	78,2	68,9	70,1	71,3
Italia	65,5	66,0	66,5	77,3	77,7	78,2	73,6	74,1	74,6	67,0	67,6	68,3