

La valutazione del valore aggiunto della scuola*

PAOLO SESTITO¹

La qualità del sistema educativo è oggi una grande questione nazionale. Vi si connettono le opportunità di crescita economica – che per un paese come l'Italia dipendono dalla capacità innovativa, non dalla presenza di risorse naturali da sfruttare, e che sarebbe controproducente perseguire tramite una sistematica compressione dei livelli salariali. Ma anche le prospettive della coesione sociale – l'Italia caratterizzandosi come un paese in cui troppo spesso la condizione professionale e reddituale dei figli ricalca quella dei genitori, con una scarsa capacità del sistema scolastico, nonostante la sua natura e il suo finanziamento quasi esclusivamente pubblici, di spezzare tale immobilità intergenerazionale.

Affrontare una simile questione significa colmare il gap medio negli apprendimenti che, crescentemente al progredire lungo il ciclo degli studi, caratterizza gli studenti italiani rispetto al dato medio Europeo o dei paesi OCSE e significa ridurre le disuguaglianze interne al sistema scolastico².

Un tratto di queste, specifico del caso italiano, è la forte differenziazione esistente tra scuole³, delle diverse regioni – con un marcato ritardo del Mezzogiorno, che si ricollega e influenza il più generale gap socio-economico di queste regioni – ma anche tra scuole della stessa regione⁴. È anche e soprattutto con riferimento a queste che nel dibattito di

* Relazione inviata al Congresso SIRD, non sottoosta a referaggio.

1 Lo scritto riprende i temi fondamentali dell'intervento svolto al VII Congresso SIRD "Università e scuola: valutare per quale società?", svoltosi a Padova nei giorni 1 - 2 - 3 dicembre 2011.

2 L'evidenza internazionale è che i due obiettivi, di innalzamento dei livelli medi e di riduzione delle disuguaglianze interne al sistema, non sono in contraddizione (cfr. ad es. Freeman et al., 2010).

3 L'obiettivo di una riduzione dei divari tra scuole e tra regioni nel caso delle scuole di primo e di secondo ciclo, differenzia queste dal segmento universitario. In quest'ultimo, una certa differenziazione della struttura dell'offerta formativa – con la concentrazione delle eccellenze in un ridotto numero di poli attivi nella competizione a livello mondiale – ha invece una sua ragione d'essere (cfr. Cipollone et al., 2012), obiettivo dell'Università non essendo quello di fornire a tutti almeno una base formativa adeguata.

4 I differenziali tra regioni (cfr., per un quadro aggiornato, Invalsi, 2012) non sembrano in prima battuta ascrivibili a una diversa disponibilità di risorse materiali: la dotazione infrastrutturale delle scuole del Mezzogiorno è di minore qualità – un dato che dipende dalle minori disponibilità degli enti locali di quell'area ma anche da una minore propensione di questi a spendere per la scuola – ma il grosso della spesa è in Italia legata alle retribuzioni del personale – in Italia finanziate dallo Stato centrale e che nel

questi ultimi anni si è fatto crescente riferimento al tema della comparabilità degli esiti formativi degli alunni tra le diverse scuole e della *accountability* delle scuole come possibile leva di miglioramento. È ai problemi connessi con l'uso di questo canale di miglioramento, e in particolare al tema del come stimare e adoperare il concetto di valore aggiunto – inteso come stima del contributo che una scuola può aver dato all'evoluzione delle competenze dei propri alunni – che sono dedicate queste pagine.

Preliminarmente è però opportuno chiarire, non foss'altro che per evitare fraintendimenti sulle opinioni di chi scrive, che l'evidenza internazionale secondo cui autonomia e *accountability* delle scuole, peraltro se e solo se abbinata l'una all'altra⁵, si associano a una migliore *performance* del sistema scolastico, non implica l'irrilevanza di altri importanti aspetti. Un innalzamento della qualità del sistema scolastico non può oggi prescindere dal puntare su docenti più qualificati e motivati – quindi meglio selezionati e meglio pagati, in sostanziale discontinuità rispetto alle prassi degli ultimi decenni, con docenti in numero elevato⁶ ma sottopagati, selezionati (ma spesso nel senso sbagliato) quasi solo dallo spauracchio del lungo precariato attraverso cui transitare e sostanzialmente abbandonati a se stessi nella loro attività quotidiana. Anche opportuno appare essere un rinnovamento degli stessi orizzonti culturali della scuola italiana – dove l'orientamento alle competenze e l'abitudine a stimolare la motivazione e l'interesse degli allievi sono spesso disattesi. Alcuni di tali temi sono legati alla *governance* del sistema scolastico, e si avrà modo di parlarne, sia pur brevemente, nelle pagine che seguono. Più in generale, però essi si ricollegano a un'esigenza di maggior attenzione sociale al tema della scuola e del suo funzionamento. E tale attenzione trascende lo stesso mondo della scuola in senso stretto: conta ad esempio un funzionamento più inclusivo del mercato del lavoro. Un mercato del lavoro inclusivo e tale da valorizzare i talenti dei più giovani che la scuola può e deve sviluppare è il miglior viatico per stimolare i ragazzi all'impegno ed evitare una pressione sociale al ribasso, al mero ottenimento del "pezzo di carta" (possibilmente con la minore fatica possibile), che tante volte caratterizza invece, specie in alcune aree del paese, il modo con cui le famiglie si relazionano con la scuola.

L'idea che gli "utenti del servizio" possano indurre una maggiore qualità delle scuole – alla stessa stregua di quanto avviene in tanti altri settori ed attività ove la pressione della domanda e le logiche di mercato "disciplinano" i comportamenti dell'offerta, è stata in Italia tradizionalmente trascurata. L'ipotesi implicita era che gli utenti – *in primis* le famiglie – non fossero in grado di formulare scelte consapevoli o che addirittura potessero formulare scelte distorte, ad esempio a favore di scuole "più facili", pronte a garantire l'ottenimento del famoso "pezzo di carta" più che allo sviluppo delle competenze; si noti che è proprio per questo motivo che quanto prima detto sul mercato del lavoro rileva anche per un buon funzionamento della scuola. Più in generale, la fornitura dei "servizi educativi" è comunque, alla stregua di tanti altri servizi professionali, un'attività i cui utenti (gli alunni, ma anche i loro genitori) per definizione hanno un *handicap* informativo, che ne rende meno ponderate

Mezzogiorno, in rapporto agli alunni, sono semmai leggermente più elevate (per la maggiore anzianità media del corpo docente di quell'area, ove più bassa è l'incidenza di docenti precari, e per via di una dimensione media delle classi che è spesso inferiore e non sempre per motivi ragionevoli, perché legati alla presenza di alunni disabili o di una struttura abitativa più frammentata e la più diffusa presenza di scuole di minore dimensione). Cfr. in generale su tali temi Cipollone e Sestito (2011).

5 Cfr. Woessmann et al. (2009)

6 Anche dopo i "tagli" degli ultimi anni, il modulo organizzativo e l'impegno orario dei docenti italiani si sostanziano in un rapporto tra studenti e docenti che, nel confronto internazionale, rimane piuttosto basso.

le scelte. Pur con queste qualificazioni, pare però indubbio che la pressione degli utenti, un fenomeno che comunque tenderà a manifestarsi, possa contribuire a innalzare la qualità della scuola. Perché ciò avvenga occorrono però meccanismi che guidino e incanalino tale spinta: regole di trasparenza – che rendano “informata” e non distorcano la pressione delle famiglie – e regole di funzionamento del sistema – che tramutino le scelte informate delle famiglie in stimoli a ben fare.

Molte di queste cose sono oggi assenti in Italia e non se ne potrà qui parlare. Vale però la pena sottolineare come un contributo rilevante stia già venendo dalla presenza delle rilevazioni Invalsi. Le fughe in avanti da taluni a volte ipotizzate⁷ sarebbero però controproducenti, anche per il rischio di enfatizzare eccessivamente il significato di tali prove – che non sono e non vogliono essere una sorta di giudizio di Dio sulle singole scuole (e tanto meno sui singoli alunni), essendo mirate a fornire innanzitutto informazioni comparate alle singole scuole – ed il cui svolgimento rischierebbe di essere deformato, più di quanto già oggi non accada, se divenissero il metro per “giudicare”, e quindi premiare o punire, le singole scuole o i singoli insegnanti. La trasparenza anche nell’uso di tali prove può però essere migliorata già da subito e in questo senso si muovono talune recenti iniziative: a partire dall’autunno 2012, con riferimento ai risultati delle rilevazioni Invalsi della precedente primavera, l’Invalsi sta ad esempio restituendo alle scuole i propri risultati al netto degli effetti stimati della presenza di *cheating* – così evitando che le scuole finiscano col farsi forti di risultati ottenuti in maniera fraudolenta⁸ – e sta prevedendo un preciso *format* dei dati che la scuola, a sua scelta, potrà rendere pubblici; tale *format* in particolare enfatizza l’evoluzione dei risultati della scuola tra i diversi gradi scolastici e il posizionamento della scuola rispetto a scuole con una popolazione studentesca simile, in termini di *background* socio-economico delle famiglie di origine, sì da approssimare il concetto di valore aggiunto. Più in generale, il documento rendicontativo standard che la bozza di regolamento sul SNV (Sistema Nazionale di Valutazione) approvato in prima lettura dal Governo nell’agosto 2012 immagina debba esser prodotto da ogni singola scuola seguendo un *format* affidato all’Invalsi, potrà col tempo includere anche altre informazioni atte a cogliere la qualità del servizio scolastico. Si potrà così evitare di enfatizzare, con eccesso spasmodico, i risultati delle rilevazioni Invalsi, che colgono aspetti importanti ma non esaustivi dei risultati raggiunti in ciascuna singola scuola.

La pressione derivante dalle “scelte di mercato” delle famiglie, oltre a non essere l’unica possibile spinta al miglioramento, non è comunque un meccanismo privo di controindicazioni e rischi. Sul piano empirico, non vi è in effetti nessun sistema scolastico governato unicamente da logiche di mercato. Il motivo è presto detto: non tutte le famiglie sono egualmente sensibili al tema qualità del servizio scolastico; svariati fattori – ad esempio la presenza di ostacoli geografici o di reti di trasporto mal funzionanti – possono limitare la reattività della domanda alla qualità del servizio, per cui quel che magari è ottenibile in un ambiente urbano non è possibile in situazioni a bassa densità di popolazione; sarebbe comunque difficile, per la sua problematicità intrinseca, avere che le scuole mal funzionanti “escano dal mercato” per venire rimpiazzate da scuole di migliore qualità. Vi è perciò il rischio che una logica di quasi-mercato esalti la dicotomia tra “scuole buone” – che, sommerse dalle richieste e spesso impossibilitate ad espandersi, razionano l’accesso dei

7 Cfr. ad esempio Abravanel (2011).

8 L’obiettivo è ovviamente quello di stimolare al controllo del *cheating*. A tale fine le informazioni sulla presenza di *cheating* nelle diverse classi della scuola sono anche fornite al dirigente scolastico e al presidente del consiglio d’Istituto (un genitore).

propri alunni, magari proprio sulla base delle loro competenze e della loro motivazione, finendo così col migliorare ulteriormente i propri risultati – e “scuole cattive” – con problemi accentuati dal dover servire quasi solo studenti difficili e poco motivati.

Oltre che regole sulla trasparenza, occorrono perciò regole che impediscano tali derive – ad esempio limitando la possibilità per le scuole di scegliersi gli studenti più bravi e motivati⁹ – e meccanismi e logiche che vadano a sovrainporsi a quelli di mercato – ad esempio, la presenza di interventi esterni che impongano, dall’alto e sulla base di un principio gerarchico–autoritativo, una riorganizzazione delle “scuole cattive”, che non possono esser lasciate andare alla deriva col loro carico di studenti innocenti. Quella di individuare situazioni critiche ove sia necessario intervenire è in effetti un’altra grande funzione d’un sistema di valutazione che compari tra loro le diverse scuole.

Una non banale precisazione è doverosa sul concetto di situazione critica. Gli esiti formativi ed educativi¹⁰ degli alunni d’una scuola possono ovviamente discendere tanto dal contesto in cui la scuola si trovi ad operare, *in primis* il background socio-economico dei propri studenti, quanto dal suo *modus operandi*. Nel linguaggio aziendale, dei risultati scadenti – e quindi la presenza di una situazione critica – possono discendere dalla scarsità delle risorse o dalla bassa produttività con cui queste vengono adoperate. Individuare come “critica” la situazione di una scuola non significa pertanto giudicare come inefficace, e quindi da rivedere, il suo *modus operandi*: ad esser necessario potrebbe essere un supplemento di risorse e un aiuto esterno, prima ancora che una riorganizzazione dei suoi processi organizzativi interni.

È proprio al fine di distinguere tra queste due possibilità che entra in campo il concetto di valore aggiunto, un costrutto analitico, di per sé non apprezzabile dagli utenti che, anche quando informati, tenderanno peraltro a reagire più agli esiti formativi ed educativi che alla quota parte di questi ascrivibile alla scuola. Visto che è all’ottenimento di elevati esiti formativi ed educativi che dovrebbe mirare un buon sistema scolastico, è importante

9 In caso di razionamento ciò significa adoperare delle vere e proprie lotterie. Più in generale ciò significa impedire che le scuole buone si caratterizzino per un costo, per le famiglie, troppo difforme da quello esistente per la generalità delle scuole (anche con riferimento ad attività accessorie, ad esempio i contributi per il laboratorio o i viaggi di istruzione). La presenza di vincoli di tale tipo non deve peraltro stupire in quello che comunque sarebbe destinato a rimanere un “mercato regolato”.

10 Esiti formativi ed educativi che possono a loro volta essere visti come l’assieme delle competenze di natura più propriamente cognitiva e dei valori e degli atteggiamenti comportamentali (i cd *skills* non cognitivi). Una “buona scuola” dovrebbe infatti consentire lo sviluppo di competenze di qualità, ma anche promuovere efficacemente i valori e le norme collettive congruenti con una società libera e democratica e preparare adeguatamente al mondo professionale e del lavoro. Trascurare nella prassi applicativa questa multidimensionalità delle misure di *outcome* rilevanti rischierebbe di distorcere i comportamenti degli attori del sistema – che saranno indotti a tenere in conto solo gli aspetti più facilmente misurabili e quindi oggetto di esplicita rilevazione, ad esempio restringendo indebitamente il curriculum alle sole cose oggetto delle rilevazioni Invalsi – e di deformare la stessa capacità segnaletica delle misure adoperate – incentivando ad esempio meccanismi di *cheating* e di vera e propria frode nell’attuazione delle prove. E’ perciò importante ragionare e costruire un sistema basato su misure intrinsecamente multidimensionali, oltre che definire procedure di contrasto di tali fenomeni discorsivi. Inoltre, proprio perché la potenza segnaletica delle rilevazioni standardizzate delle competenze è molto elevata – sul piano dell’indicazione anche culturale di quali siano le competenze chiave verso cui le scelte educative debbano andare – tali strumenti vanno adoperati con cautela e consapevolezza, evitando di dare indicazioni incongrue e precipitose al sistema. Pur attesa l’importanza di tali questioni, riassumibili nella considerazione secondo cui l’*outcome* di una scuola non possa ridursi ai risultati dei suoi alunni nelle prove Invalsi, da qui in poi tale questione verrà trascurata, implicitamente immaginando di basare sulle rilevazioni Invalsi la misura dell’*outcome* di una scuola da cui poi derivare il concetto di valore aggiunto.

individuare le situazioni scolastiche caratterizzate da esiti insoddisfacenti; è però solo ragionando in termini di valore aggiunto che si potrà concludere, con riferimento al singolo caso concreto, se il problema richieda un'azione di supporto – maggiori risorse – o di riorganizzazione – un modo più efficace di adoperare le risorse esistenti.

Beninteso, all'atto pratico in molti casi potrà essere necessario sia supportare che riorganizzare. Inoltre, anche ove la conclusione fosse a favore della necessità di maggiori risorse, la concreta definizione del “che fare” richiede di ragionare sul come intervenire. L'unico caso in cui ci si potrebbe limitare a ragionare sul valore aggiunto è quello in cui tale metrica venga adoperata per premiare le scuole ritenute migliori, agli stimoli derivanti da tali forme di premialità¹¹ venendo affidata la funzione di indurre miglioramenti nella qualità delle diverse scuole. Ma l'uso di forme premiali non supera i problemi prima visti per i meccanismi di quasi mercato: non è possibile far progressivamente mancare le risorse alle “scuole cattive”, che non possono essere lasciate andare alla deriva, privandole di risorse, col loro carico di innocenti studenti. Nel caso della scuola, prima o poi (e meglio prima che poi) occorrerà intervenire, non essendo contemplato e contemplabile il “fallimento di mercato” delle singole scuole¹²!

Se l'obiettivo non è (solo) quello di premiare le singole scuole o di cercare di meglio informare le scelte delle famiglie¹³, la valutazione non può però limitarsi a stimare, dall'esterno, il valore aggiunto delle diverse scuole – rendendone comparabili i risultati e soprattutto tenendo conto del fatto che questi possono dipendere dalle conoscenze pregresse degli alunni di una certa scuola, dal loro *background* familiare e da altri fattori di contesto che poco hanno a che fare con le scelte compiute nelle singole scuole. Occorre anche passare per un esame *in loco* del *modus operandi* della singola scuola, per poter individuare su cosa e come concretamente intervenire. Una valutazione che voglia innescare processi di miglioramento deve perciò anche comprendere la dinamica di funzionamento di quella singola organizzazione scolastica, sì da individuare le specifiche ragioni dei risultati ottenuti e orientare le stesse azioni innovative. Per fare tutto ciò non basta misurare gli esiti formativi ed educativi degli alunni di quella scuola e neppure la loro evoluzione nel tempo

11 Se lo stimolo deve venire dalla presenza di un possibile premio, i meccanismi di elargizione dello stesso – e quindi il cosa si intenda premiare – dovrebbero essere ben segnalati in anticipo. A contare dovrebbe essere l'anticipazione d'un possibile premio – potenzialmente ottenibile da tutti, impegnandosi adeguatamente – e non tanto la gratificazione, per pochi, *ex post*. La potenza d'un sistema premiale – che ove tarato sul valore aggiunto e non sugli esiti in livello dei propri alunni potrebbe stimolare a ben fare tutti e non solo i pochi le cui condizioni di partenza consentano di raggiungere un livello assoluto abbastanza elevato – non è però priva di rischi. Dovendo precisare prima cosa si intenda premiare, forte è il rischio di incentivare le scuole a occuparsi solo degli aspetti esplicitamente considerati nella “gara”, col rischio di corrompere la stessa validità degli indicatori adoperati nella “gara”. In altre forme, ritorna il problema generale già prima discusso con riferimento al rischio di sovraccaricare di significati le rilevazioni Invalsi. Sulla questione, sia pure con riferimento ad un diverso contesto, si veda anche Barley e Neal (2012), che discutono di come strutturare una “gara”, esplicitamente definendola in termini relativi (*rank tournament*) e basandola sull'uso di misure che, non mirando a comparazioni nel tempo ma solo tra “partecipanti alla gara”, siano non prevedibili e quindi non tali da distorcere i comportamenti dei partecipanti.

12 Questa obiezione all'uso della premialità vale peraltro per le scuole, ma non per i loro dirigenti, per i quali premi e sanzioni ben potrebbero essere adoperati senza particolari rischi (beninteso, al di là di quelli, non piccoli, connessi con la qualità degli indicatori adoperati per misurarne la *performance*, a partire da quella della scuola a ciascuno di essi affidata, e delle possibili distorsioni che l'esplicitazione di un meccanismo premiale può indurre nei loro comportamenti).

13 Nel senso di cercare di indurle a ragionare in termini di valore aggiunto e non solo di confronto tra livelli assoluti.

o stimare il valore aggiunto attribuibile alla scuola: è necessario interpretare, *in loco*, come certi esiti derivino dall'insieme dei processi interni alla scuola e delle risorse a sua disposizione, tenendo conto del contesto in cui essa opera.

Vi è qui in altri termini uno spazio che è proprio anche dei processi di autovalutazione. Questa è proficua nella misura in cui – proprio perché svolta da chi dispone di quella *soft information* che dall'esterno è più difficile rilevare e codificare – riesce a individuare concrete piste di revisione dell'organizzazione interna. Essa è perciò innanzitutto un processo di continua riconsiderazione, a fini di miglioramento, del proprio *modus operandi*.

Così interpretata l'autovalutazione non va peraltro contrapposta alla valutazione esterna. Momenti di valutazione esterna – la presenza di una funzione ispettiva – e, più in generale, l'individuazione di criteri e strumenti comuni e di standard omogenei per la stessa autovalutazione, servono innanzitutto ad impedire che, banalmente quanto inutilmente, ogni scuola (ogni *team* di insegnanti o ogni singolo insegnante) dia peso solo agli aspetti di rendicontazione e, soprattutto, li deformi attribuendosi il “massimo dei voti”. I riferimenti esterni comuni e omogenei possono però servire anche ad orientare ciascuna scuola verso una prospettiva sistemica di analisi organizzativa interna, utile a riconoscerne le componenti essenziali e le reciproche relazioni ed a leggere il singolo caso comparativamente agli altri. L'autovalutazione, pur avendo il proprio vantaggio comparato proprio nell'uso di informazioni di tipo *soft* e non codificate, ha senso solo se praticata con criteri omogenei e adoperando un comune modello interpretativo¹⁴.

Avendo chiarito che stimare e identificare il valore aggiunto non è quindi la fine della storia, occorre ora comunque chiedersi più in dettaglio cosa sia questo costrutto del valore aggiunto. La via maestra – che empiricamente è stata in Italia sperimentata all'interno del progetto VSQ e che da questo anno scolastico potrà essere generalizzata, quantomeno con riferimento al momento di passaggio dalla V primaria alla I media inferiore – è quella di seguire l'evoluzione degli apprendimenti nel tempo per ogni alunno, considerando poi la media per scuola (o per gruppi di insegnanti) di questi “progressi cognitivi” al posto del livello medio delle competenze.

Affinamenti ulteriori dovrebbero in realtà tenere conto del fatto che tale evoluzione può avere una sua dinamica complessa. Quando si considerano gli apprendimenti relativi a una singola disciplina bisognerebbe ad esempio tener conto della presenza di effetti incrociati, per cui le competenze iniziali in una disciplina possono favorire lo sviluppo delle competenze in un'altra disciplina e non solo quelle della propria disciplina. L'impatto del livello iniziale delle competenze in un dato anno di studi può inoltre riverberarsi non solo sui livelli raggiunti durante l'anno immediatamente successivo, ma anche avere effetti a più

14 I processi di autovalutazione rappresentano inoltre un possibile anello di congiunzione tra il livello di scuola e quanto accade al suo interno. Un maturo processo di autovalutazione all'interno delle singole scuole può infatti consentire di fornire un *feedback* anche ai singoli insegnanti, o comunque ai diversi *team* di insegnanti, basato non esclusivamente su indicatori desumibili dall'esterno – ad esempio i risultati (o l'evoluzione dei risultati) nelle rilevazioni Invalsi dei loro alunni. Si noti che oggi in Italia il collegamento tra singoli (*team* di) insegnanti e rilevazioni Invalsi non è istituibile a livello centralizzato e non è quindi certo dall'esterno della singola scuola che si possa valutare, su una base comparativa e standardizzata, i singoli insegnanti. Ma l'impossibilità tecnologica e organizzativa non è l'unico aspetto rilevante; vi sono anche aspetti di inopportunità da considerare. Basare solo su dati di questo tipo la valutazione dei singoli insegnanti accentuerebbe i rischi di distorsioni nei dati medesimi. Soprattutto, basare esclusivamente su dati di questo tipo la valutazione dei singoli insegnanti trascurerebbe l'informazione di tipo *soft* che è invece disponibile all'interno della scuola e renderebbe impossibile fornire un *feedback* ai valutati sugli aspetti ove sia possibile migliorare.

lungo termine (ad esempio, la performance negli studi universitari potrebbe dipendere dalle competenze nel dominare la lingua acquisite nel ciclo primario, prima ancora che dalle competenze acquisite nel ciclo secondario superiore). La dinamica di ciascun singolo anno potrebbe inoltre risentire di fenomeni di *mean reversion*, legati anche al momento del ciclo di vita in cui ciascun alunno si viene a trovare (pur a parità di percorsi a lungo termine, vi può essere chi sviluppa prima e chi dopo certe competenze). Anche a poter disporre di informazioni longitudinali sui singoli alunni – cosa che come detto solo ora inizia ad essere possibile e con riferimento solo ad uno specifico momento di passaggio, da un singolo grado scolastico a quello immediatamente successivo¹⁵ – la stima del valore aggiunto richiederebbe perciò un’attenta considerazione della possibile complessità della dinamica sottostante. L’obiettivo delle stime del valore aggiunto dovrebbe del resto essere quello di cogliere il contributo della scuola alle prospettive di sviluppo di più lungo termine dei propri alunni, evitando di “premiare” performance e risultati di breve termine.

Si è detto del fatto che gli esiti formativi sono un costrutto multidimensionale, solo malamente circoscrivibile nei risultati delle rilevazioni Invalsi e del fatto che, anche accettando in prima battuta una simile limitazione, vi sono problemi intrinseci nell’uso di tali rilevazioni come misura dell’*output* di una certa scuola. Vi è da chiedersi in che misura l’uso di tali misure in un’ottica di valore aggiunto, inteso come stima del progresso nel tempo di tali misure e non del loro livello, esalti o attenui tali problemi. La buona notizia è che è poco plausibile ritenere che venga aumentato il peso delle distorsioni derivanti dalla specifica natura delle rilevazioni di base. Visto che la *performance* degli studenti in una certa prova è funzione della motivazione personale e degli incentivi a ben fare¹⁶ oltre che delle loro competenze, la considerazione dei progressi cognitivi anziché del livello delle competenze dovrebbe semmai ridurre il peso di questi aspetti motivazionali, in generale abbastanza costanti nel tempo per ciascun individuo, e di risposta agli incentivi¹⁷ come componente esplicativa della variabilità complessiva¹⁸. La cattiva notizia è che l’ordinamento delle scuole in base al valore aggiunto è plausibilmente più variabile nel tempo di quello ottenibile sulla base dei risultati in livello¹⁹. In parte ciò è una buona cosa, nel senso che si rimuovono quei fattori di contesto e di composizione che costantemente differenziano le diverse scuole e da cui il concetto di valore aggiunto vuole affrancarsi. È proprio in tale modo che si può evitare di dire che una scuola frequentata da ragazzi con uno sfavorevole

15 Col tempo, questa possibilità è però destinata ad accrescersi: le rilevazioni Invalsi – che non sono un’esauritiva misura degli esiti formativi ed educativi degli studenti italiani ma hanno il pregio di essere comparabili tra le diverse scuole – consentiranno di ricostruire l’intero percorso scolastico e formativo, dall’inizio della scuola primaria al termine del secondo ciclo (quando l’Invalsi realizzerà rilevazioni, come previsto dalla normativa, anche nella V classe della secondaria superiore).

16 Questi possono essere influenzati dall’eventuale natura *high stakes* di una data prova, ad esempio laddove questa sia parte di un esame ufficiale.

17 Almeno laddove le due prove abbiano la stessa natura e non si confrontino i risultati per lo stesso individuo di prove *high stakes* e *low stakes*.

18 Nel caso di meccanismi premiali, focalizzare questi sul valore aggiunto, inteso come sviluppo nel tempo delle competenze, oltre a fornire stimoli diffusi ed evitare che solo chi sia abbastanza vicino al target da raggiungere abbia un incentivo a darsi da fare, potrebbe anche aiutare nel contrasto del *cheating*: gli insegnanti di un dato grado scolastico sono i più naturali “controllori” dei comportamenti, in sede di prova, degli studenti (e degli insegnanti) del grado immediatamente inferiore (cfr. anche Ferrer, 2012).

19 L’esperienza mentale che qui si sta effettuando è quello di considerare tre rilevazioni sulle competenze – coi conseguenti tre possibili ordinamenti delle scuole ciascuno basato su una delle tre rilevazioni – e le due stime di valore aggiunto ottenibili dal confronto nel tempo delle prime due e delle seconde due rilevazioni.

background socio-economico è una scuola poco efficace!. In parte però questa variabilità nel tempo dipende dal fatto che la considerazione della performance in termini di progressi cognitivi, anziché di livello delle competenze, esalta il peso, come componente della variabilità complessiva, degli errori di misurazione comunque presenti in ciascuna singola rilevazione²⁰: una scuola potrebbe apparire come particolarmente efficace (inefficace) solo perché nella rilevazione finale (iniziale) determinati fattori casuali abbiano favorevolmente influenzato la momentanea *performance* dei suoi studenti.

Come detto, al momento è solo con riferimento a pochi casi che si può seguire la *performance* nel tempo dei singoli studenti. Occorre quindi parlare brevemente anche di quelle che sono delle possibili *proxies* del valore aggiunto basate su dati che non siano propriamente longitudinali a livello di singolo studente. In altri termini, cosa si può fare laddove si disponga solo dei dati di diverse rilevazioni sezionali, magari ripetute nel tempo ma senza che sia possibile misurare i progressi cognitivi dei singoli studenti tra queste diverse rilevazioni sezionali?

Posto che l'obiettivo è quello di comparare le scuole al netto dei fattori di composizione e di contesto che le distinguono l'una dall'altra, una prima via è quella di considerare i risultati in livello relativi a una singola rilevazione (ad esempio la rilevazione di un certo anno riferita alla V primaria) ma depurando dagli effetti stimati di composizione della popolazione studentesca della singola scuola (metodo A). Una seconda via è quella di confrontare i risultati ottenuti dalle classi di due diversi gradi della stessa scuola (metodo B) ed una terza considera infine le differenze nel tempo ma calcolate al livello non dei singoli studenti ma della scuola nel suo complesso (metodo C).

E' facile vedere pregi e difetti di tali metodi. Il primo è poco affidabile laddove a contare siano soprattutto fattori di contesto e di composizione che risultino essere non osservabili (ad esempio per indisponibilità di molte misure sui fattori di composizione della popolazione degli alunni) ed il cui effetto non possa quindi essere stimato ed eliminato tramite opportune tecniche statistiche (ad esempio tramite l'uso di modelli di regressione multipla). Nel caso italiano, il *background* familiare del singolo alunno è ricostruito con una certa precisione nella V primaria, nella I media inferiore e nella II superiore, mentre si dispone solo di poche informazioni nella II primaria (perché i ragazzi censiti sono troppo giovani per rispondere in maniera affidabile a quesiti del genere) e nella III media inferiore (nel cui caso la prova è parte dell'esame di stato²¹). Gli altri due metodi offrono specularmente il vantaggio di poter tenere conto anche di quei fattori di composizione e di contesto che differenziano tra loro le diverse scuole e che siano non osservabili: questi verrebbero semplicemente rimossi, pur senza passare per la stima dei loro effetti. Tale vantaggio però si concretizza se e solo se tali fattori non cambiano nel tempo. Confrontare la V primaria e la II primaria di oggi di una data scuola può ben approssimare il confronto tra gli alunni oggi di V e i risultati di quegli stessi alunni in II se gli odierni alunni di II sono molto simili – in base a caratteristiche osservabili e non – agli alunni di II di tre anni prima. Per gli stessi

20 Ogni singola rilevazione per definizione sarà imprecisa a causa di errori di misurazione o di accadimenti del tutto estemporanei e poco legati alle competenze sottostanti dei diversi alunni: la *performance* in un dato giorno e in una data prova di un singolo studente, o anche di un'intera classe, potrà ad esempio discendere dallo stato di salute, dal clima o dall'umore in quel momento prevalenti. Se tali errori di misurazione, come probabile, sono indipendenti tra loro nel tempo, la considerazione delle variazioni nel tempo tra due rilevazioni porta ad amplificarne il peso come componente esplicativa della variabilità complessiva.

21 Per inciso, ciò rende tra l'altro *high stakes* tale prova, rendendola meno comparabile, anche a livello di singolo studente, con le altre.

motivi, confrontare la V primaria di oggi con la V primaria di un anno prima può dare indicazioni sulla presenza di tendenze al miglioramento o al peggioramento d'una data scuola – un dato estremamente rilevante al fine di monitorare gli effetti dell'arrivo di un nuovo dirigente scolastico o l'efficacia di un certo piano di miglioramento adottato dalla scuola – ma se e solo se la composizione della V primaria di oggi e di un anno prima non sono mutate; se tra gli studenti di oggi, o tra quelli di un anno prima, vi sono studenti “anomali” (nel senso di avere risultati anomalmente positivi o negativi) o se i fattori di contesto della scuola sono molto cambiati, attribuire il miglioramento o il peggioramento ai comportamenti della scuola (al nuovo dirigente scolastico o all'implementazione di un dato piano di interventi) diventa impossibile.

I metodi B e C sono inoltre poco robusti laddove le politiche di passaggio da un grado scolastico al successivo variano rispettivamente tra scuole o nel tempo (per la stessa scuola): ove varino tra scuole, un elevato differenziale cognitivo tra alunni, ad esempio, di III e di I media inferiore, potrà essere legato al fatto che quella scuola seleziona grandemente i propri studenti nel percorso sino alla III media, per cui arriveranno a tale traguardo solo un gruppo fortemente selezionato della propria popolazione iniziale di studenti²²; anche ipotizzando una costanza nel tempo della composizione degli studenti iscritti nella I classe della media inferiore, il gruppo di quanti siano arrivati nell'anno corrente in III media inferiore potrebbe differire dal gruppo di quanti vi siano arrivati l'anno prima se nel tempo sono variati i criteri di selezione adoperati dalla singola scuola.

Come detto, il metodo C apparentemente dovrebbe consentire di focalizzarsi sulla tendenza al miglioramento o al peggioramento di una data scuola e quindi sul contributo al margine proveniente dalla presenza di un nuovo dirigente scolastico o dall'attuazione di un piano di miglioramento appositamente formulato, grandezze di indubbia rilevanza dal punto di vista della *governance* del sistema nel suo complesso. Esso è però particolarmente soggetto all'impatto della già prima richiamata presenza di errori di misurazione. Qualsiasi elemento temporaneo – la presenza ad esempio di pochi alunni che, in una data annualità, siano particolarmente brillanti o problematici – influenzerà la stima del valore aggiunto basata su tale metodo²³.

Alla luce di tali considerazioni, pare ragionevole evitare di dare troppo affidamento alle variazioni nel tempo della *performance* di ciascuna singola scuola. Per tener conto della presenza di fattori temporanei – *lato sensu* assimilabili agli errori di misurazione – una buona prassi potrebbe esser anzi quella di escludere le osservazioni “estreme”, concentrandosi su indicatori rappresentativi della situazione di scuola diversi dalla media²⁴. Quanto agli altri due metodi, una buona prassi potrebbe essere invece quella di combinarli tra loro. Ad esempio, all'interno del metodo A, in aggiunta a tutti i fattori di composizione osservabili, si potrebbero includere come ulteriore fattore di controllo i risultati medi nei gradi

22 Questo meccanismo è particolarmente rilevante nel caso della scuola media superiore.

23 In effetti Falzetti e Ricci (2012) rilevano come le variazioni da un anno all'altro nell'ordinamento delle scuole in base al valore aggiunto, anche dopo aver tenuto conto degli effetti dei fattori di composizione lungo le linee del metodo A, sono piuttosto marcate.

24 La media di scuola – costruita sommando i risultati di tutti i singoli alunni e dividendo per la numerosità – è fortemente influenzata dalla presenza di osservazioni singole aberranti; indicatori come la mediana (che identifica il risultato dello studente che divide esattamente in due metà l'intera distribuzione dei risultati individuali) o una media troncata (ovvero una media costruita dopo aver eliminato le osservazioni con valori estremamente bassi o estremamente alti) potrebbero meglio indicare quale sia la tendenza “centrale”.

precedenti misurati a livello di scuola²⁵; oppure, all'interno del metodo B si potrebbero adoperare dati prima depurati dall'effetto dei fattori di composizione che siano comunque osservabili²⁶.

Implicitamente è a una simile combinazione delle logiche di cui ai metodi A e B che fa riferimento il *frame* disposto dall'Invalsi nel piano di restituzione dei risultati nelle rilevazioni Invalsi alle scuole messo in atto dall'autunno del 2012: oltre al risultato grezzo, della scuola e di vari possibili benchmark di rilievo (la regione e l'Italia nel suo complesso), si evidenzia la differenza tra i risultati di scuola e quelli di una scuola con simile composizione della popolazione studentesca (la logica del metodo A) e si evidenzia il posizionamento – sempre rispetto a questi diversi benchmark – dei diversi gradi scolastici presenti in quella scuola e coperti dalle rilevazioni Invalsi (la logica del metodo B, in quanto tale e messo in atto su dati già depurati per gli effetti di composizione, ovvero sia una combinazione dei metodi A e B).

Riferimenti bibliografici

- Abravanel R. (2011). Test pubblici per scuole migliori. *Meritocrazia (Blog)*, <http://www.corriere.it>.
- Barlevy G., N. Derek (2012). Pay for Percentile. *American Economic Review*.
- Cipollone P., Sestito P. (2011). *Il capitale umano*. Bologna: il Mulino.
- Cipollone P., Montanaro P., Sestito P. (2012). *Il capitale umano per la crescita economica: possibili percorsi di miglioramento del sistema d'istruzione in Italia*. QEF 122, Banca d'Italia.
- Falzetti P., Ricci R. (2012). *L'evoluzione nel tempo della performance delle singole scuole come misura del loro valore aggiunto*. Mimeo, lavoro presentato al workshop “Banca d'Italia – INVALSI: il valore aggiunto delle scuole italiane: come misurarlo?”. Roma, 31 gennaio 2012.
- Frontini F., Montanaro P. (2012). *Il valore aggiunto della scuola come confronto tra classi di diverso grado: misure per la scuola primaria*. Mimeo, lavoro presentato al workshop “Banca d'Italia – INVALSI: il valore aggiunto delle scuole italiane: come misurarlo?”. Roma, 31 gennaio 2012.
- Ferrer E. G. (2012). Ecco perché c'è chi bara nei test Invalsi. *La voce info* (http://www.lavoce.info/articoli/-scuola_universita/pagina1003137.html).
- Freeman R.B., Machin S., Viarengo M. (2010). *Variation in Educational Outcomes and Policies across Countries and of Schools within Countries*. NBER Working Papers 16293.
- INVALSI (2012). *Rilevazioni nazionali sugli apprendimenti 2011-12. Il quadro di sistema*. (http://www.governo.it/GovernoInforma/Dossier/valutazione_istruzione_istruzione_istruzione/rapporto_rilevazione_apprendimenti_2012.pdf)
- Lüdemann E., Schütz G., Woessmann L., West M.R. (2009). School Accountability, Autonomy and Choice Around the World. *Journal of Economics*, 99, 3, pp. 283-285.
- Martini A., L. Marcucci (2012). *Dalla performance degli studenti al contributo della scuola: il ruolo dei fattori di composizione nei singoli segmenti del sistema scolastico, mimeo, lavoro presentato al workshop*. “Banca d'Italia – INVALSI: il valore aggiunto delle scuole italiane: come misurarlo?”. Roma, 31 gennaio 2012.

25 Martini et al. (2012) considera ad esempio un modello, per la V primaria, che utilizza l'indicatore di background familiare (costruito con precisione in tale grado scolastico) ma anche, a livello di scuola, i risultati odierni nelle classi II, il valore aggiunto essendo definito come il residuo stimato di un simile modello.

26 Questa è la strada adoperata da Frontini e Montanaro (2012).