

Uno studio longitudinale sul valore aggiunto come misura di efficacia scolastica: risultati ed elementi di problematicità

Alessandra Rosa - Università di degli Studi di Bologna - alessandra.rosa3@unibo.it
Liliana Silva - Università degli Studi di Bologna - liliana.silva@unibo.it

A longitudinal study on value-added indicators for measuring school effectiveness: results and critical aspects

Il dibattito degli ultimi anni sul valore aggiunto come indicatore di efficacia scolastica sottolinea l'esigenza di disegni di rilevazione e analisi dei dati di tipo longitudinale. Alla luce di tali orientamenti, il presente studio ha preso in esame una leva di oltre 700 studenti frequentanti 12 scuole secondarie di I grado della provincia di Bologna, seguendola dal momento dell'ingresso a quello dell'uscita dalla scuola media. Attraverso la comparazione dei dati emersi nel triennio, i risultati della ricerca confermano alcune problematiche di ordine metodologico evidenziate dalla letteratura internazionale, fornendo inoltre interessanti spunti di riflessione sull'uso dei modelli di valore aggiunto in un momento in cui, anche nel nostro Paese, essi sembrano godere di crescente considerazione nell'ambito delle procedure finalizzate alla valutazione di sistema.

Parole chiave: efficacia scolastica, valore aggiunto, accountability e miglioramento, studio longitudinale, scuola secondaria di I grado, comprensione del testo.

The ongoing debate about value-added models as indicators of school effectiveness highlights as a primary need the detection and analysis of longitudinal data. Consequently, the present study followed from the first to the last year a cohort of more than 700 students from 12 junior high schools in the province of Bologna. At present, value-added models are increasingly taken into account even in our country, aiming to the evaluation of the school system. Through the comparison of data obtained during the three years, results of the present research confirm some of the methodological issues already discussed in the international literature, while also providing interesting insights on the use of value-added.

Keywords: school effectiveness, value-added models, accountability and school improvement, longitudinal study, junior high school, reading comprehension skills.

169

ricerche

Benché il contributo sia frutto del lavoro congiunto delle due autrici, **Alessandra Rosa** ha scritto i §§ 1, 3 e 4, **Liliana Silva** ha scritto i §§ 2, 5 e 6.

Uno studio longitudinale sul valore aggiunto come misura di efficacia scolastica: risultati ed elementi di problematicità

1. Caratteristiche e definizioni dei modelli di valore aggiunto



L'esigenza di una scuola equa ed efficace, che si pone in modo sempre più pressante nello scenario attuale della società della conoscenza e del *lifelong e lifewide learning*, evidenzia la portata culturale e politica della problematica presa in esame nel presente contributo, relativa alla misurazione dell'efficacia scolastica non solo in un'ottica di responsabilità e trasparenza nei confronti dell'opinione pubblica, ma anche nella prospettiva di un miglioramento continuo dell'offerta formativa e dei processi di insegnamento/apprendimento. La ricerca di modelli e procedure che consentano di pervenire a misure valide e attendibili dell'efficacia dei sistemi di istruzione e delle singole scuole risulta pertanto fondamentale da un lato per la cosiddetta *accountability*, ovvero per la rendicontazione esterna dell'operato degli istituti e degli insegnanti, divenuta centrale anche a seguito dell'accresciuta autonomia ad essi assegnata a livello gestionale, organizzativo e didattico; dall'altro lato per incentivare processi di valutazione/autovalutazione che, a partire dalla riflessione e dal confronto sui risultati emersi, consentano di definire interventi di miglioramento in relazione ai punti deboli riscontrati.

Sulla base delle esigenze delineate, nell'ambito di quel filone della ricerca educativa internazionale solitamente definito come *School Effectiveness Research* (SER) è emersa e si affermata, soprattutto negli ultimi due decenni, la proposta di modelli basati sul *valore aggiunto*. Tale concetto, le cui origini vanno ricercate nel settore dell'economia e della produzione di beni materiali, è stato trasposto al settore dell'istruzione e proposto da vari studiosi afferenti al campo di indagine come possibile soluzione al problema di individuare indicatori di efficacia più affidabili e validi di quelli tradizionalmente utilizzati per la valutazione dei sistemi scolastici. Questi ultimi infatti, essendo basati sui "punteggi grezzi" degli studenti in prove standardizzate di profitto somministrate al termine di determinati gradi/cicli scolastici, sono ritenuti indicativi non tanto dell'efficacia degli istituti frequentati e dell'istruzione ricevuta, quanto piuttosto delle differenze nella composizione della popolazione scolastica in termini di *background* socio-culturale: è noto infatti che tra rendimento scolastico e ambiente familiare di provenienza vi è una forte associazione e che gli alunni in condizione di svantaggio ottengono, mediamente, risultati inferiori rispetto agli studenti di status più elevato.

In che modo si differenziano, rispetto a tali misure, gli indicatori ottenibili mediante modelli di valore aggiunto? È possibile rispondere a tale domanda partendo dalla descrizione che ne fornisce l'OCSE in un rapporto dedicato alla tematica dal titolo *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*, in cui essi vengono definiti come una categoria di modelli statistici che stimano il contributo delle scuole ai progressi degli studenti verso il

raggiungimento di obiettivi educativi stabiliti o prescritti “al netto” di altri fattori che influenzano tali progressi (OECD, 2008). In altre parole, il valore aggiunto indica l'accrescimento nei livelli di apprendimento degli studenti specificamente riconducibile ai processi di istruzione, in quanto i risultati che essi ottengono nelle prove di profitto vengono elaborati statisticamente per tener conto del peso esercitato da variabili antecedenti quali il retroterra socio-culturale di provenienza e la loro “dotazione iniziale” in termini di conoscenze/competenze possedute, che costituiscono potenti predittori della riuscita scolastica.

A questo proposito, è necessario sottolineare il fatto che, nell'ambito della *School Effectiveness Research*, l'espressione “valore aggiunto” è stata utilizzata per indicare almeno tre diversi approcci alla misurazione dell'efficacia scolastica (Schagen, Hutchison, 2003): a) misurazione del rendimento attraverso un'unica rilevazione, “depurando” però i punteggi grezzi conseguiti dall'influenza del *background* familiare degli studenti; b) misurazione del progresso attraverso più rilevazioni effettuate a distanza di tempo, “depurando” i punteggi grezzi conseguiti al termine di determinati periodi di istruzione dall'influenza riconducibile al rendimento pregresso degli studenti; c) misurazione del progresso attraverso più rilevazioni effettuate a distanza di tempo, “depurando” i punteggi grezzi conseguiti al termine di determinati periodi di istruzione dall'influenza sia del rendimento pregresso, sia del *background* familiare degli studenti.

Il concetto di valore aggiunto ha dunque assunto accezioni e significati differenti ma, come testimonia ad esempio il rapporto dell'OCSE precedentemente citato, la letteratura più recente sembra ampiamente concorde nel circoscrivere l'uso di tale espressione soltanto al terzo degli approcci menzionati. In definitiva, il valore aggiunto, in quanto misura di *progresso* piuttosto che di *status*, si configura come differenza tra un risultato *osservato* – il punteggio effettivamente ottenuto dagli studenti nelle prove standardizzate – e un risultato *atteso* – determinato dalle caratteristiche di sfondo e di partenza degli studenti; se la differenza assume segno positivo si può realmente parlare di valore aggiunto, mentre in caso di segno negativo sarebbe forse più opportuno parlare di “valore sottratto”. La tecnica statistica tipicamente impiegata per determinare tale differenza è rappresentata dall'analisi della *regressione*, che consente di individuare e “isolare” – nell'ambito dell'analisi della varianza derivata dalla regressione – la parte di varianza associata ai risultati degli studenti che, non essendo attribuibile ai predittori inseriti nel modello, può considerarsi, con le dovute cautele, riconducibile all'effetto della scuola e dei processi di istruzione.

I risultati delle analisi di valore aggiunto variano a seconda del modello adottato e la presentazione degli stessi può essere strutturata in base ai livelli di aggregazione dei dati utilizzati come riferimento per il calcolo (singolo studente, classe/insegnante, scuola, area locale o regionale), nonché in base agli usi e ai destinatari previsti. A tale riguardo, il rapporto dell'OCSE individua tre principali tipologie di impiego dei dati ottenuti mediante modelli di valore aggiunto, ciascuna delle quali implica il riferimento, almeno in via prioritaria, a diversi tipi di stakeholder: la prima è finalizzata a supportare l'*accountability* scolastica, la seconda le scelte degli studenti e delle loro famiglie e infine la terza il miglioramento degli istituti e dei processi educativi (OECD, 2008).

I primi due obiettivi individuati hanno entrambi a che fare con l'istanza, richiamata in apertura del contributo, di una rendicontazione sociale dell'operato delle scuole mediante la pubblicazione dei risultati raggiunti dagli studenti, realizzando sistemi informativi in grado di fornire validi e attendibili indicatori di efficacia atti a supportare processi decisionali a vari livelli (dalle scelte di politica



scolastica e allocazione delle risorse a quelle delle famiglie). Il terzo si riferisce invece alla necessità, sempre più sottolineata negli ultimi anni (ad es. Giovannini, Tordi, 2009), di un uso “interno” dei dati provenienti da valutazioni esterne, in cui le informazioni prodotte a livello di sistema sono messe al servizio dei singoli istituti per contribuire ad alimentare processi di riflessione e autovalutazione che, a loro volta, portino a definire piani di miglioramento.

In quest’ottica perde importanza la creazione di “classifiche” delle scuole e degli insegnanti, rintracciabile ad esempio nella tradizione delle *League Tables* inglesi, mentre acquisisce valore la ricaduta formativa dei risultati, il loro porsi quali dati cui dirigenti scolastici e docenti possono fare riferimento per una migliore comprensione/regolazione del proprio agire professionale e quali informazioni utili a supportare il monitoraggio e il miglioramento continuo dei processi di istruzione.

2. Il dibattito relativo alle indagini longitudinali sul valore aggiunto



Il dibattito relativo alle finalità e alla definizione sempre più accurata dei modelli per l’analisi del valore aggiunto ha trovato ampio respiro all’interno delle diverse esperienze sviluppatesi in ambito internazionale. All’interno di questo quadro, una delle tematiche centrali su cui vari studiosi hanno richiamato l’attenzione è indubbiamente rappresentata dal tipo di disegno necessario nell’impostazione delle procedure di rilevazione/analisi dei dati e, in particolare, dall’importanza di adottare una prospettiva longitudinale, che in riferimento al valore aggiunto si presenta come necessaria, ma al tempo stesso problematica. Tale esigenza si ricollega non solo all’intento di pervenire a stime più corrette degli effetti complessivamente rilevabili a livello di istituto e di classe, ma anche a quello di sondare la stabilità nel tempo degli indicatori di valore aggiunto relativi alle singole scuole/classi.

Per quanto alcuni autori abbiano teorizzato e avanzato ipotesi relative alle possibilità di un modello di valore aggiunto longitudinale (ad es. Gray *et al.*, 1996; Luyten, 1994), la prospettiva a lungo termine è stata oggetto di ricerche empiriche solo negli ultimi anni (cfr. ad es. Goldhaber, Hansen, 2008). A seguito delle difficoltà riscontrate nella misurazione nel corso di un solo anno (ad es. Giovannini, Tordi, 2009) si è infatti ritenuto necessario ampliare il *framework* temporale della ricerca, per poter considerare i progressi degli studenti all’interno di un arco temporale più ampio.

Come è stato messo in evidenza nel precedente paragrafo, per il calcolo degli indicatori di valore aggiunto sono necessarie almeno due rilevazioni dei dati sugli stessi soggetti; tuttavia, alcuni studiosi hanno indicato il triennio come l’arco temporale “minimo” per analisi che consentano di verificare la stabilità nel tempo delle misure effettuate (Kyriakydes, Creemers, 2008; Thomas *et al.*, 2007; Van de Grift, 2009). L’ampliamento del *framework* temporale considerato permette infatti di ridurre gli errori di misura (Rothman, 2010) e di potenziare l’*effetto formativo dei dati*, valorizzando il monitoraggio continuo e quindi la possibilità di ottenere informazioni importanti per attuare miglioramenti negli istituti e nelle classi (Creemers, Kyriakydes, 2006).

Come sopra accennato, gli studi longitudinali offrono vantaggi dal punto di vista dell’analisi dei dati, ma anche difficoltà e problematiche che, soprattutto se non opportunamente esplicitate e affrontate, possono indebolire la validità e l’at-

tendibilità dei risultati ottenuti. Innanzitutto è necessario disporre di un sistema di codifica/archiviazione dei dati che consenta di seguire l'allievo per un periodo di tempo molto lungo, nel quale inevitabilmente saranno presenti variazioni (Amrein-Beardsley, 2008); occorre inoltre prevedere l'insorgere di comportamenti opportunistici da parte di alcuni soggetti interessati (ad es. effetto *gaming*) e curare il processo di informazione sui metodi e sui fini, che deve accompagnare la costruzione di un impianto per molti aspetti complesso e delicato (Ricci, 2008); infine è necessario considerare che le variazioni nel tempo sono il frutto di dinamiche complesse, non riconducibili esclusivamente all'effetto della scuola e alle prassi educative correnti (effetti ritardati, effetti *life cycle*, effetti incrociati tra ambiti disciplinari ecc).

La dimensione longitudinale è inoltre strettamente collegata alla problematica dei dati mancanti: più misure sono necessarie per un calcolo accurato, più aumenta la possibilità di avere una mancanza di dati per il calcolo stesso, generando instabilità e incertezza nella restituzione di "classifiche" e risultati alle scuole (ad es. Ding, 2009; Van de Grift, 2009). Anche per i modelli di valore aggiunto, l'aspetto dei *missing data* fa sostanzialmente riferimento ai soggetti che, per vari motivi (trasferimenti, ritiri, bocciature, assenze ecc.), presentano dati incompleti e vengono esclusi dalle analisi; il problema riguarda inoltre gli studenti che non sono in grado di sostenere la prova (certificati, stranieri non alfabetizzati ecc.), i quali tuttavia incidono sull'efficacia complessiva della didattica. Pur costituendo uno dei principali nodi problematici messi in luce da vari studiosi, la letteratura internazionale non sembra offrire una definizione precisa della possibile soluzione. Nell'ambito dei sistemi implementati in altri Paesi per la rilevazione e l'analisi del valore aggiunto il problema è stato affrontato in diversi modi. Il TVAAS (*Tennessee Value Added Assessment System*) sviluppato da William Sanders, ad esempio, tratta i dati relativi ai soggetti mancanti equiparandoli alla media rilevata a livello di distretto ma, come sottolinea Kupermintz (2003), ciò può produrre distorsioni soprattutto per le classi caratterizzate da un elevato numero di dati mancanti. Portela *et al.* (2013), invece, propongono la definizione di un indice per la valutazione delle evoluzioni delle performance scolastiche strutturato sulle misure di valore aggiunto e sulle buone pratiche sviluppatesi nel tempo. Inserire l'impatto statistico dei dati mancanti nei modelli di valore aggiunto è certamente possibile, ma se non fatto con precisione e analitica definizione degli indici il risultato finale, ancora una volta, può generare informazioni scarsamente attendibili (Wiley, 2006).

La scelta più ricorrente in letteratura sembra essere quella di escludere i soggetti con dati mancanti dalle analisi di valore aggiunto, basata in sostanza sul presupposto che essi siano distribuiti in modo casuale tra le varie scuole/classi. Van de Grift (2009), tuttavia, sottolinea come anche questa scelta sia discutibile, in quanto la distribuzione dei dati mancanti può essere considerata casuale soltanto in parte: i tassi di mobilità, di assenteismo, di ripetenza o di abbandono, ad esempio, non sono i medesimi in tutti gli istituti, ma tendono ad essere particolarmente elevati soprattutto nelle scuole caratterizzate da un'utenza di basso livello socio-culturale e da indicatori di risultato sotto la media.



3. Obiettivi e impostazione della ricerca

Ponendosi in continuità con gli studi sul valore aggiunto già effettuati nell'ambito del Dottorato di ricerca in Pedagogia sperimentale della Sapienza - Università di Roma¹, ma al tempo stesso differenziandosi da essi non solo per il livello scolastico considerato ma soprattutto per la scelta innovativa di adottare un disegno di ricerca longitudinale, nel triennio compreso tra gli anni scolastici 2008/2009 e 2010/2011 è stato realizzato uno studio nella scuola secondaria di I grado che ha coinvolto e collegato tra loro tre ricerche di dottorato², ciascuna incentrata su uno dei tre anni del progetto ma con un progressivo ampliamento del raggio di analisi dei dati. In particolare, nell'indagine svolta nel secondo anno è stato possibile prendere in esame un arco temporale di due anni, confrontando i risultati degli allievi con quelli ottenuti, dagli stessi allievi, nell'ambito dell'indagine realizzata nel primo anno; allo stesso modo, nell'indagine effettuata nel terzo anno è stato possibile riflettere sull'andamento dei risultati relativi all'intero triennio considerato.

A partire dal quadro teorico delineato nei precedenti paragrafi e sulla scorta degli studi effettuati in altri Paesi, il progetto di ricerca ha inteso esplorare nel contesto italiano potenzialità e limiti degli indicatori di valore aggiunto per la misurazione dell'efficacia scolastica. Le ipotesi formulate e sottoposte a verifica – incentrate sul confronto tra tali indicatori e quelli basati sui punteggi grezzi, sulla comparazione tra diversi livelli di analisi (*scuola e classe*) e sull'esame dell'andamento delle misure effettuate nell'arco di tempo in esame – hanno mirato, nel complesso, a fornire utili spunti di riflessione in merito alla seguente controversa questione: è possibile utilizzare il valore aggiunto per valutare – ed eventualmente premiare/sanzionare – le scuole e gli insegnanti?

Il campione coinvolto nella ricerca, tratto da una popolazione di riferimento rappresentata da tutte le scuole secondarie di I grado attive sul territorio della provincia di Bologna nell'a.s. 2008/2009, comprende 12 istituti per un totale di 36 classi (tre per ciascun istituto) e oltre 700 studenti. La selezione delle unità di analisi da tale popolazione è avvenuta mediante campionamento *per obiettivi* o *di giudizio* o *ragionato*, espressioni con cui si indica quel “disegno che ha come criterio il giudizio del ricercatore stesso, che sceglie le unità di analisi proprio sulla base della sua aspettativa di ottenere risposte che meglio si confanno alle finalità dell'indagine” (Lucisano, Salerno 2002, p. 138). Tale piano di campionamento, quindi, pur essendo di tipo non probabilistico e limitando di conseguenza la possibilità di generalizzare i risultati ottenuti, è stato scelto in base alle finalità e alle caratteristiche del progetto di ricerca, rispetto alle quali risultava necessario includere nello studio scuole rispondenti ai seguenti due requisiti di fondo: la collocazione in zone e contesti diversificati, allo scopo di coinvolgere istituti caratterizzati da un'utenza eterogenea in termini di *background* socio-culturale; la disponibilità di



- 1 Nell'ambito del Dottorato in Pedagogia sperimentale la tematica della misura dell'efficacia scolastica per mezzo del valore aggiunto è stata precedentemente affrontata dal dott. Cristiano Corsini (XX ciclo) e dalla dott.ssa Claudia Tordi (XXI ciclo), le cui ricerche hanno preso in esame l'ultimo anno della scuola primaria facendo riferimento, rispettivamente, alle competenze degli alunni in lettura e in matematica.
- 2 Le tre indagini coinvolte nel progetto di ricerca longitudinale - coordinate e supervisionate dalla Prof.ssa Maria Lucia Giovannini e dal Prof. Pietro Lucisano - sono state realizzate dalla dott.ssa Margherita Ghetti (XXII ciclo), dalla dott.ssa Alessandra Rosa (XXIII ciclo) e dalla dott.ssa Liliana Silva (XXV ciclo).

almeno tre classi/sezioni parallele, per consentire l'analisi della varianza *intra-scuola* delle misure di efficacia.

La scelta dell'ambito entro cui rilevare gli apprendimenti degli studenti – da cui partire per elaborare indicatori e graduatorie di efficacia relativi alle scuole e alle classi partecipanti alla ricerca – ha privilegiato le abilità di lettura e comprensione dei testi, che indubbiamente occupano un posto di primo piano tra le abilità di base che la scuola ha il compito di fornire agli studenti: esse non solo sono di natura trasversale e dunque necessarie per affrontare con successo lo studio di qualunque disciplina scolastica, ma risultano anche indispensabili per la vita al di fuori della scuola in una società che sempre più richiede capacità di ampliamento/aggiornamento continuo delle conoscenze possedute. In ciascuno dei tre anni considerati si è dunque proceduto alla messa a punto, alla taratura e alla somministrazione di prove strutturate di comprensione del testo, utilizzando item di ancoraggio tra una prova e l'altra e sottoponendo ogni volta i risultati ad item analisi mediante procedure riconducibili al modello dell'*Item Response Theory* (IRT), che consente di incrementare la precisione e l'attendibilità delle misure mettendo in relazione la difficoltà dei quesiti utilizzati e l'abilità dei soggetti rispondenti. Come illustra la Tabella 1 sotto riportata, la medesima leva di studenti è stata coinvolta in un ciclo di quattro rilevazioni successive, consentendo di analizzare in prospettiva diacronica l'andamento degli indicatori di valore aggiunto nel triennio considerato.



Classe prima - a.s. 2008/09		Classe seconda a.s. 2009/10	Classe terza - a.s. 2010/11
I somministrazione <i>Ottobre 2008</i>	II somministrazione <i>Maggio 2009</i>	III somministrazione <i>Aprile/Maggio 2010</i>	IV somministrazione <i>Aprile/Maggio 2011</i>
Prova di comprensione del testo <i>entrata prima media</i>	Prova di comprensione del testo <i>uscita prima media</i>	Prova di comprensione del testo <i>uscita seconda media</i>	Prova di comprensione del testo <i>uscita terza media</i>

Tab. 1: La scansione temporale dello studio longitudinale nella scuola secondaria di I grado

Per raccogliere i dati relativi al *background* familiare degli studenti, indispensabili per elaborare le misure di valore aggiunto, ma anche al fine di rilevare ulteriori informazioni di contesto e di processo utili ad approfondire e interpretare tali misure – focalizzando in particolare l'attenzione sulle caratteristiche, sulle percezioni e sulle prassi degli insegnanti coinvolti –, sono stati inoltre messi a punto e utilizzati dei questionari rivolti tanto agli studenti quanto ai loro docenti di italiano.

4. Principali risultati emersi dal primo biennio dello studio

Rifacendosi alle tecniche statistiche già utilizzate nel primo anno dello studio longitudinale, gli indicatori di valore aggiunto relativi al secondo anno sono stati ottenuti mediante un'equazione di *regressione lineare multivariata* in cui, come *variabile dipendente*, è stato considerato il punteggio nella prova di comprensione del testo svolta dagli studenti in uscita dalla seconda media, mentre come *predittori* o *regressori* sono stati inseriti: *a)* il *rendimento progressivo*, ovvero i punteggi nelle due prove svolte in prima³; *b)* l'*indice socio-culturale familiare* (ISC), costruito me-

3 I coefficienti di correlazione osservati tra la variabile dipendente e il rendimento pre-

dante analisi fattoriale a partire da alcune variabili di sfondo relative all'ambiente di provenienza degli studenti (tra cui il titolo di studio e l'occupazione dei genitori)⁴. Nel complesso, i regressori inseriti nel modello di analisi contribuiscono a "spiegare" una percentuale di varianza nei risultati degli studenti pari al 79%, che ne conferma l'elevato potere predittivo sul rendimento scolastico.

Come già osservato nel primo anno dello studio e come ipotizzato in base agli esiti emersi da altre indagini effettuate in contesti diversi e/o a differenti livelli del sistema di istruzione (ad es. Choi et al., 2004; Corsini, 2009; Giovannini, Tordi, 2009), l'uso di indicatori di valore aggiunto (VA) in luogo delle tradizionali misure basate sui punteggi grezzi (PG) tende a modificare notevolmente il quadro che emerge in termini di efficacia: il confronto tra i due tipi di indicatori evidenzia infatti la presenza sia di scuole/classi risultate efficaci sulla base dei PG che, in termini di VA, mostrano invece indicatori di segno negativo, sia di scuole/classi efficaci in base agli indicatori di VA che mostrano invece, in termini di PG, valori inferiori alla media.

Anche la comparazione tra i cosiddetti "effetto-scuola" ed "effetto-classe" – misurati in termini di *varianza spiegata* nei punteggi degli studenti a parità di altre condizioni rilevanti (rendimento pregresso e *background* socio-culturale) – ha rivelato risultati in linea con quanto emerge dalla letteratura internazionale (ad es. Bressoux, 1995; Luyten, 2003) e dalle precedenti ricerche sul valore aggiunto svolte nell'ambito del Dottorato in Pedagogia sperimentale (cfr. note 1 e 2): mentre nel primo caso si registra una percentuale pari al 5%, nel secondo essa sale al 18% (cfr. Tabella 2). La classe sembra dunque configurarsi come dimensione di analisi più rilevante e significativa rispetto a quella rappresentata dalla scuola nel suo insieme: lo scarto osservato tra le quote di varianza spiegata dai due livelli di aggregazione dei dati, insieme alla sostanziale disomogeneità riscontrata tra gli indicatori di VA relativi alle diverse classi appartenenti a ciascuna scuola, induce infatti a concludere che a fare realmente la differenza nei progressi degli studenti sia la specifica classe frequentata, più che l'istituto di appartenenza.



Tabella ANOVA – Varianza VA tra SCUOLE							
		Somma dei quadrati	df	Media dei quadrati	F	Sig.	Quota di varianza spiegata
VA* Scuola	Fra gruppi	26,399	11	2,400	2,488	,005	5,0%
	Entro gruppi	502,601	521	,965			
	Totale	529,000	532				
Tabella ANOVA – Varianza VA tra CLASSI							
		Somma dei quadrati	df	Media dei quadrati	F	Sig.	Quota di varianza spiegata
VA* Classe	Fra gruppi	96,468	35	2,756	3,167	,000	18,2%
	Entro gruppi	432,532	497	,870			
	Totale	529,000	532				

Tab. 2: Effetto-scuola ed effetto-classe

gresso sono risultati elevati – e significativi al livello 0,01 – per entrambe le prove svolte in ingresso e in uscita dalla prima: 0.84 nel primo caso e 0.85 nel secondo.

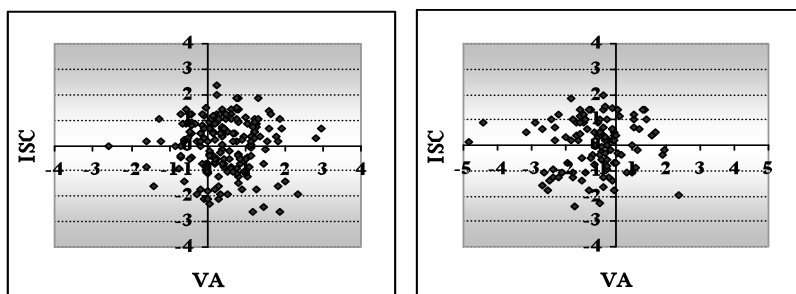
- 4 La correlazione tra l'indice e la variabile dipendente è risultata elevata – e significativa al livello 0,01 – tanto a livello di scuola (0.81) quanto a livello di classe (0.73).

Un ulteriore confronto effettuato, reso possibile dall'adozione di un disegno longitudinale e dunque dal collegamento con l'indagine svolta nel primo anno dello studio, è quello relativo alla comparazione tra gli indicatori di VA ottenuti al termine della prima media e quelli osservati al termine della seconda. Confermando i risultati di alcuni studi longitudinali condotti in altri Paesi (ad es. Aaronson *et al.*, 2007; Bressoux, Bianco, 2004; Goldhaber, Hansen, 2008; Gorard *et al.*, 2013; Hill *et al.*, 2010; Thomas *et al.*, 2007), gli esiti relativi al primo biennio della nostra ricerca mettono in luce una forte discontinuità nel tempo delle misure di valore aggiunto relative alle varie scuole e soprattutto alle singole classi, anche nei casi di permanenza dello stesso insegnante nei due anni scolastici considerati: mentre infatti i punteggi grezzi appaiono generalmente piuttosto stabili, permanendo nel biennio superiori o inferiori alla media in 10 scuole su 12 e in 31 classi su 36, gli indicatori di VA mostrano nella maggior parte dei casi notevoli slittamenti – verso l'alto o verso il basso – nel passaggio dal primo al secondo anno (si vedano, a questo proposito, gli esempi riportati nel successivo Grafico 5). Questo andamento complessivamente “fluttuante” degli indicatori di VA induce a riflettere non solo sull'affidabilità e validità delle misure di efficacia, ma anche sull'uso dei dati per esempio per classificare e premiare – come viene fatto in altri Paesi – le scuole o gli insegnanti: come valutare, ad esempio, un docente che, in riferimento ai medesimi studenti, risulti efficace nella classe prima e inefficace in seconda?

Un ultimo risultato su cui vorremmo soffermarci, anche in relazione a quanto precedentemente affermato circa l'esigenza di coniugare efficacia ed equità, riguarda la “distribuzione” del valore aggiunto tra gli studenti: ci sono differenze tra gruppi di allievi di diverso livello socio-culturale? Nel Grafico 1 sotto riportato – relativo alle classi più e meno efficaci, cioè collocate agli estremi della graduatoria ottenuta al termine della seconda media – si illustra il modo in cui si distribuiscono gli allievi considerando i valori assunti dalle variabili VA (asse delle x) e ISC (asse delle y). Osservando il primo diagramma si nota che gli studenti tendono a concentrarsi nella metà destra in quanto si fa riferimento alle classi con indicatori di VA positivi, ma sembrano ripartirsi in maniera piuttosto equilibrata tra la parte superiore e quella inferiore; nel secondo diagramma si osserva una situazione speculare a quella appena descritta: in questo caso, infatti, gli studenti si concentrano nella metà sinistra in quanto si fa riferimento alle classi con indicatori di VA negativi, ma tendono a ripartirsi tra la parte superiore e quella inferiore in maniera altrettanto equilibrata.

La distribuzione del valore aggiunto tra gli studenti appare dunque sostanzialmente “equa”: in altre parole, nelle classi in cui si aggiunge o si sottrae valore, ciò avviene per tutti gli studenti, senza disparità connesse al *background* familiare. Nella maggior parte dei casi, inoltre, le classi con VA positivo mostrano una concomitante riduzione delle differenze di rendimento tra gli studenti, con un decremento del valore assunto dal *coefficiente di variazione* tra la prima e la seconda media, mentre nelle classi con VA negativo il coefficiente al contrario aumenta.





Graf. 1: Confronto tra la distribuzione degli alunni nelle classi più efficaci (VA positivo) e meno efficaci (VA negativo)

Approfondendo ulteriormente l'analisi e articolando l'indice ISC in 5 livelli (invece di limitarsi a considerarlo come superiore o inferiore alla media) emergono tuttavia differenze, all'interno del campione preso in esame, tra gli studenti appartenenti alle modalità estreme: come mostra la seguente Tabella 3, ottiene valore aggiunto positivo il 45,7% degli studenti di livello socio-culturale basso, mentre la percentuale sale al 61,9% per quelli di livello alto; viceversa, solo il 33% degli allievi di livello socio-culturale alto ottiene valore aggiunto negativo, mentre la percentuale sale al 43,5% in relazione a quelli di livello basso. Tali dati assumono particolare rilievo se confrontati con quelli riportati nella successiva Tabella 4, in cui la distribuzione degli studenti in base alla variabile ISC viene incrociata con quella delle misure di valore aggiunto ottenute, invece che al termine della seconda media, in uscita dalla prima. Comparando le due tabelle si nota infatti un evidente "allargamento della forbice" nel corso del biennio in esame: le differenze tra gli studenti di livello socio-culturale alto e basso, che alla fine della prima media appaiono di lieve entità, tendono invece a diventare più consistenti al termine del successivo anno scolastico. Ciò induce a supporre, pur in presenza di una distribuzione del valore aggiunto complessivamente equa, che le differenze tra gli studenti più vantaggiati e svantaggiati in termini di *background* familiare tendano comunque a emergere e ad accentuarsi nel corso del tempo⁵.



		Indice socio-culturale (ISC)					Totale
		Basso	Medio-basso	Medio	Medio-alto	Alto	
Valore aggiunto II media	Negativo	43,5%	38,9%	47,7%	41,1%	33,3%	42,6%
	Nulla	10,9%	11,9%	11,6%	10,7%	4,8%	11,1%
	Positivo	45,7%	49,2%	40,7%	48,2%	61,9%	46,3%
Totale		100% (N=46)	100% (N=126)	100% (N=172)	100% (N=168)	100% (N=21)	100% (N=533)

Tab. 3: Distribuzione degli alunni in base all'indice ISC e agli indicatori di VA: uscita seconda media

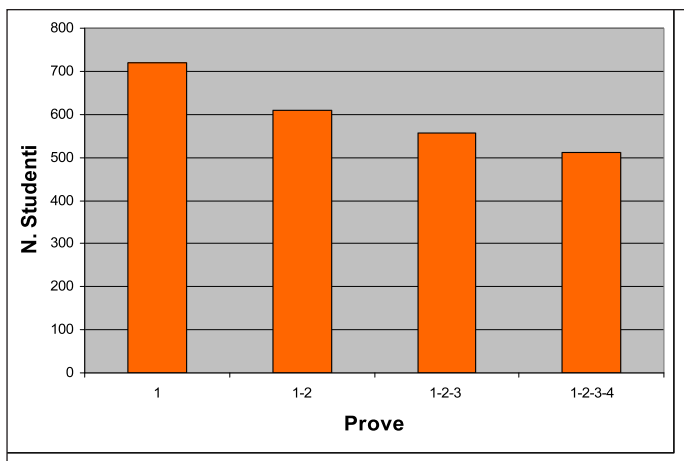
5 Per una presentazione più ampia e approfondita dei risultati emersi al termine del primo biennio dello studio si rimanda al volume di Rosa A. (2013).

		Indice socio-culturale (ISC)					Totale
		Basso	Medio-basso	Medio	Medio-alto	Alto	
Valore aggiunto I media	Negativo	41,3%	42,9%	41,3%	47,0%	38,1%	43,3%
	Nulla	8,7%	7,9%	8,1%	10,7%	9,5%	9,0%
	Positivo	50,0%	49,2%	50,6%	42,3%	52,4%	47,7%
	Totale	100% (N=46)	100% (N=126)	100% (N=172)	100% (N=168)	100% (N=21)	100% (N=533)

Tab. 4: Distribuzione degli alunni in base all'indice ISC e agli indicatori di VA: uscita prima media

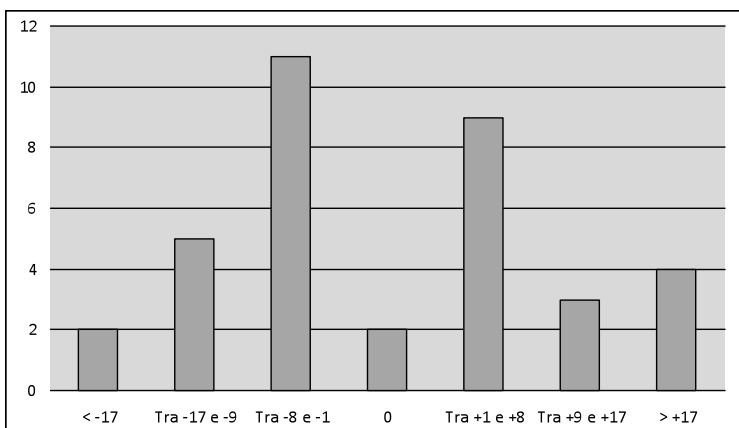
5. Principali risultati al termine del triennio della ricerca

Le analisi dei dati svolte al termine del terzo e ultimo anno della ricerca hanno seguito le procedure già adottate nei primi due anni, brevemente descritte nel precedente paragrafo, facendo riferimento ai dati ottenuti per mezzo della somministrazione delle prove di comprensione del testo e dei questionari agli studenti appartenenti al campione. È sin da ora importante sottolineare come, se all'inizio dell'indagine il campione era composto da oltre 700 studenti, al termine del terzo anno coloro che hanno sostenuto la prova durante tutte e quattro le somministrazioni e quindi inclusi nel calcolo degli indicatori di valore aggiunto sono risultati in numero nettamente inferiore (cfr. Grafico 2). La scelta fatta durante la definizione iniziale del disegno longitudinale è stata infatti quella di escludere gli studenti che non avessero completato tutte le prove della ricerca.



Graf. 2: Soggetti validi per le analisi di valore aggiunto nelle quattro somministrazioni

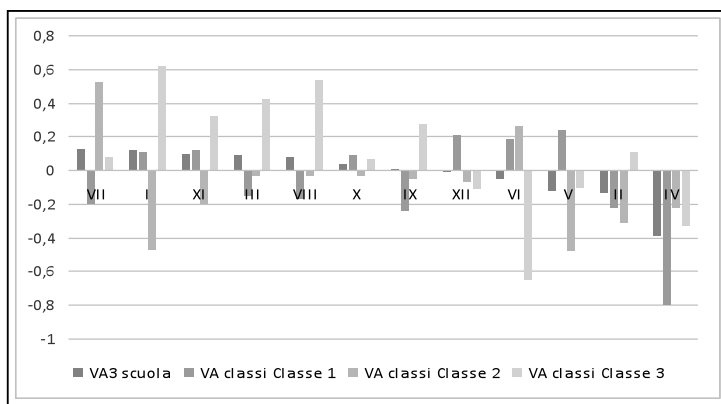
Rispetto ai risultati ottenuti al termine del secondo anno, la differenza tra i risultati ottenuti mediante la misura dei punteggi grezzi e degli indicatori di valore aggiunto risulta emergere nuovamente: se consideriamo le graduatorie generate per mezzo delle due misure, le differenze nelle posizioni ottenute dalle classi risultano essere quelle riportate nel seguente Grafico 3.



Graf. 3: Differenze nelle posizioni delle classi tra le graduatorie basate sui punteggi grezzi e sugli indicatori di VA



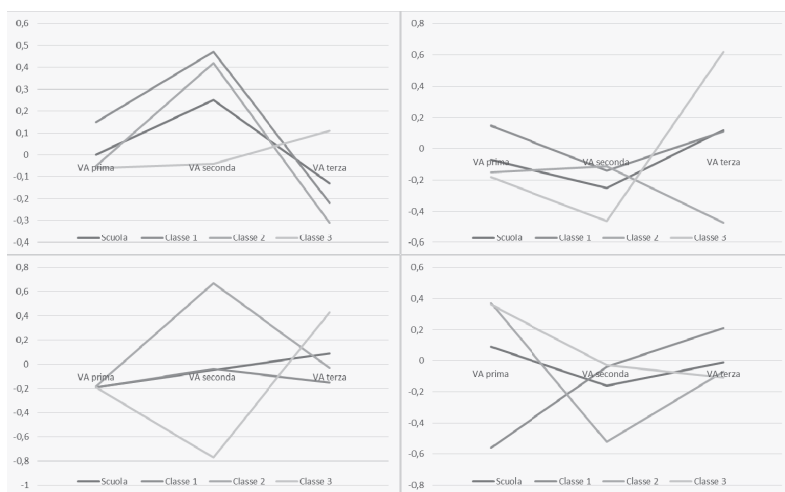
Per quanto riguarda invece l'analisi della varianza tra scuole e classi, considerando un peggioramento della significatività dei risultati ottenuti rispetto al secondo anno della ricerca, è possibile evincere come i progressi misurati per mezzo degli indicatori di valore aggiunto risultino variare tra le classi in misura maggiore che non tra le scuole: ad una quota di varianza spiegata dall'“effetto scuola” dell'1,8%, corrisponde infatti una quota di varianza spiegata dall'“effetto classe” dell'8,6%. Tale varianza è apprezzabile nel Grafico 4, dove al dato relativo agli indicatori di valore aggiunto misurati al termine del triennio per ogni scuola è affiancato quello delle rispettive classi, al fine di rilevare facilmente le importanti differenze tra le stesse classi all'interno delle medesima scuola.



Graf. 4: Indicatori di valore aggiunto ottenuti dalle scuole e dalle rispettive classi durante il terzo anno dell'indagine

Il terzo aspetto considerato, l'andamento longitudinale delle rilevazioni per mezzo del valore aggiunto, ci permette di considerare la differenza più apprezzabile rispetto ai dati relativi all'anno precedente: se al termine del secondo anno era stato riscontrato un aumento dell'“effetto classe” e dell'“effetto scuola” rispetto al primo anno della ricerca, al termine del triennio tale aspetto risulta essere inferiore

rispetto a quello del secondo anno. Inoltre, gli indicatori di valore aggiunto presentano una forte disomogeneità e quindi instabilità nel triennio in esame, soprattutto tra le classi della stessa scuola, come è possibile desumere dal seguente Grafico 5 dove sono riportati, a titolo di esempio, gli andamenti nei tre anni di quattro delle scuole comprese nel campione.



Graf. 5: Andamento degli indicatori di valore aggiunto osservati nel triennio in 4 scuole coinvolte nell'indagine e nelle rispettive 3 classi

Un'ultima considerazione riguarda invece la distribuzione del valore aggiunto tra gli studenti in relazione ai livelli socio-culturali di appartenenza: al termine del terzo anno è possibile confermare quanto già emerso alla fine del secondo (cfr. Grafico 1) e quindi una distribuzione del valore aggiunto in cui, nel complesso, non si evidenziano sostanziali differenze connesse al *background* degli allievi. Si rileva tuttavia, rispetto al secondo anno, una situazione diversa in merito alle differenze tra gli studenti di livello socio-culturale rispettivamente *alto* e *basso*, che sembrano tornare contenute come già messo in evidenza al termine del primo anno della ricerca.

Pur trattandosi di un risultato interessante, in base al quale si potrebbe supporre che la scuola si ponga come agente di equità nei confronti degli studenti riducendo gli effetti dell'indicatore ISC, riteniamo necessari ulteriori approfondimenti ed elementi di conoscenza al fine di supportare tale interpretazione.

Considerazioni conclusive

Dal presente studio longitudinale emerge come il valore aggiunto rappresenti un modello di misura dell'efficacia scolastica certamente più preciso rispetto alla misura ottenuta per mezzo dei punteggi grezzi. Emergono tuttavia anche diverse problematiche che ci permettono di mettere in discussione l'uso di questo modello come realmente valido ed equo, tra cui in particolare l'instabilità dei risultati relativi al valore aggiunto in prospettiva longitudinale. Se la letteratura evidenzia la necessità di effettuare le rilevazioni nell'arco di almeno un triennio, alcune criticità sono emerse dalla presente ricerca con riferimento alla possibilità di una misura

che richieda somministrazioni per un arco di tempo prolungato. È stato infatti rilevato, innanzitutto, il grande problema della “mortalità” dei soggetti del campione: se all’inizio dell’indagine questo poteva comprendere oltre 700 studenti, al termine del triennio, a causa delle assenze in una o più somministrazioni, dei trasferimenti e delle bocciature, il campione si è drasticamente ridimensionato (cfr. Grafico 2). È stato inoltre sottolineato l’andamento disomogeneo degli indicatori di VA relativi alle classi nel corso del triennio, confermando quindi l’instabilità dei dati nel lungo periodo.

In definitiva, la prospettiva longitudinale appare fondamentale per la definizione di un modello di efficacia scolastica strutturato sui progressi degli studenti, che necessariamente deve riferirsi a più anni; occorre tuttavia considerare le problematiche che il modello comporta, per predisporre nuove prospettive sin dalla fase di definizione del disegno stesso della ricerca.

È utile inoltre considerare l’incidenza della composizione del campione degli studenti nel triennio: sin dal principio si è infatti scelto di escludere dall’analisi gli allievi certificati e gli stranieri non alfabetizzati. Risulta tuttavia necessario considerare come la misura dell’efficacia a livello di classe non possa trascurare la presenza effettiva degli alunni suddetti nelle classi stesse, così come non può non tener conto della variabilità delle caratteristiche di alcune categorie nei tre anni (come nel caso degli stranieri non alfabetizzati) (cfr. Giovannini, 2012).

I limiti presentati ci permettono, pertanto, di mettere in discussione l’uso del modello del valore aggiunto nell’ambito della misura dell’efficacia di scuole e classi; in particolare, risulta pericoloso applicare la logica che ne prevede l’uso per la presentazione di graduatorie e la distribuzione di premi determinati in relazione a tali misure (Gorard et al., 2013). L’uso del modello del valore aggiunto sarebbe quindi utilizzabile perché produttore di utili *feedback*, ma non come unico indicatore (Scherrer, 2011).

L’analisi dei risultati di uno studio longitudinale sul valore aggiunto assume così anche una valenza politica e sociale: l’obiettivo non è infatti il raggiungimento del miglior artificio statistico, ma quello di permettere che la prospettiva meramente rendicontativa sia effettivamente accompagnata da quella migliorativa e che possa essere reale strumento di sostegno per la scuola, perché riesca a garantire le conoscenze necessarie al cittadino della società della conoscenza in un’ottica di equità.

In conclusione è dunque possibile affermare che le criticità messe in luce dallo studio effettuato richiedono ulteriori indagini e approfondimenti, che coinvolgano anche aspetti ed elementi qui non presi in considerazione riguardanti, ad esempio, il confronto tra ambiti disciplinari diversi o l’esplorazione delle variabili relative al contesto classe associate agli indicatori di valore aggiunto positivi e negativi. Ciò è necessario non solo per dare un ulteriore contributo al dibattito scientifico sull’argomento, ma anche al fine di supportare l’interpretazione e l’uso delle misure di efficacia scolastica mediante dati e informazioni che aiutino a comprenderne meglio il senso e la portata.



Riferimenti bibliografici

- Aaronson D., Barrow L., Sanders W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), pp. 95-135.
- Amrein-Beardsley A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), pp. 65-75.
- Bressoux P. (1995). Les effets du contexte scolaire sur les acquisitions des élèves: effet-école et effets-classes en lecture. *Revue Française de Sociologie*, 36(2), pp. 273-294.
- Bressoux P., Bianco M. (2004). Long-term teacher effects on pupils' learning gains. *Oxford Review of Education*, 30(3), pp. 327-345.
- Choi K., Seltzer M., Herman J., Yamashiro K. (2004). *Children left behind in AYP and Non-AYP schools: Using student progress and the distribution of student gains to validate AYP*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (ERIC Document Reproduction Service No. ED483404).
- Corsini C. (2009). *Il valore aggiunto in educazione. Un'indagine nella scuola primaria*. Roma: Nuova Cultura.
- Creemers B.P.M., Kyriakides L. (2006). Critical analysis of the current approaches to modeling educational effectiveness: the importance of establishing a dynamic model. *School Effectiveness and School Improvement*, 17(3), pp. 347-366.
- Ding C.S. (2009). Measurement issues in designing and implementing longitudinal evaluation studies. *Educational Assessment, Evaluation and Accountability*, 21(2), pp. 155-171.
- Giovannini M.L. (2012). Valore aggiunto ed efficacia delle scuole. *Rivista dell'Istruzione*, 1/2, pp. 41-48.
- Giovannini M.L., Tordi C. (2009). Misura del valore aggiunto e miglioramento dell'insegnamento. Riflessioni da un'indagine empirica nelle scuole primarie bolognesi. In G. Domenici, R. Semeraro (Eds.), *Le nuove sfide della ricerca didattica tra saperi, comunità sociali e culture*. Roma: Monolite.
- Goldhaber D., Hansen M. (2008). *Is it just a bad class? Assessing the stability of measured teacher performance*. Working Paper n. 5, CRPE, University of Washington Bothell.
- Gorard S., Hordosy R., Siddiqui, N. (2013). How unstable are "school effects" assessed by a value-added technique? *International Education Studies*, 6(1), pp. 1-9.
- Gray J., Goldstein H., Jesson D. (1996). *Changes and improvements in schools' effectiveness: trends over five years*, *Research Papers in Education*, 11(1), pp. 35-51.
- Hill H.C., Umland K., Ring Kapitula L. (2010). *Validating value-added scores: an instructional analysis*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Denver, CO, May 2010.
- Kyriakides L., Creemers B.P.M. (2008). A longitudinal study on the stability over time of school and teacher effects on student outcomes. *Oxford Review of Education*, 34(5), pp. 521-545.
- Kupermintz H. (2003). Teacher effects and teacher effectiveness: a validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis*, 25(3), pp. 287-298.
- Lucisano P., Salerni A. (2002). *Metodologia della ricerca in educazione e formazione*. Roma: Carocci.
- Luyten H. (1994). Stability of school effects in Dutch secondary education: the impact of variance across subjects and years. *International Journal of Educational Research*, 21(2), pp. 197-216.



- Luyten H. (2003). The size of school effects compared to teacher effects: an overview of the research literature. *School Effectiveness and School Improvement*, 14(1), pp. 31-51.
- OECD (2008). *Measuring improvements in learning outcomes. Best practices to assess the value-added of schools*. Paris: OECD Publications Service.
- Portela M.C., Camanho A.S., Keshvari A. (2013). Assessing the evolution of school performance and value-added: trends over four years. *Journal of Productivity Analysis*, 39, pp. 1-14.
- Ricci R. (2008). *La misurazione del valore aggiunto nella scuola*. Fondazione Giovanni Agnelli, Working Paper n. 9, 12.
- Rosa A. (2013). *Il valore aggiunto come misura di efficacia scolastica. Un'indagine empirica nella scuola secondaria di I grado*. Roma: Nuova Cultura.
- Rosa A., Giovannini M.L. (2013). *Valore aggiunto ed efficacia delle scuole e degli insegnanti: risultati di un'indagine empirica*. In corso di pubblicazione.
- Rothman R. (2010). Beyond Test Scores: Adding Value to Assessment. *School Administrator*, 67(2), pp. 20-26.
- Schagen I., Hutchison D. (2003). Adding value in educational research: the marriage of data and analytical power. *British Educational Research Journal*, 29(5), pp. 749-765.
- Scherrer J. (2011). Measuring teaching using value-added modelling: the imperfect panacea. *NASSP Bulletin*, 95(2), pp. 122-140.
- Thomas S., Peng W.J., Gray J. (2007). Modelling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, 33(3), pp. 261-295.
- Van de Grift W. (2009). Reliability and validity in measuring the value added of schools. *School Effectiveness and School Improvement*, 20(2), pp. 269-285.
- Wiley E.W. (2006). *A Practitioner's Guide to Value Added Assessment*. Tempe, Arizona State University.